DRDC Toronto number: CR-2002- 018

Assessing the Impact of Multi-Sensor Data Fusion on Command and Control Operations in the HALIFAX Class Frigate:

Recommendations for Measures of Performance and Detailed Test Plan.

by:
M.L. Matthews, R.D.G. Webb, A.R. Keeble

Human*systems* Incorporated®
111 Farquhar St., 2nd floor
Guelph, ON N1H 3N4

Project Manager:
Kim Iwasa-Madge
(519) 836 5911

PWGSC Contract No. W7711-7-7404/01-SV
Call-Up 7404-21

On behalf of
DEPARTMENT OF NATIONAL DEFENCE

as represented by
Defence R&D Canada – Toronto
1133 Sheppard Avenue West
North York, Ontario, Canada
M3M 3B9

DRDC - Toronto Scientific Authority:
Carol McCann
(416) 635-2190

February 2002

Executive Summary

This report provides a comprehensive Test and Evaluation (T&E) plan for evaluating the impact of future command decision aid technologies (COMDAT) on human and operational performance in the Halifax class operations room. The plan details an incremental approach to T&E that will generate reliable and valid performance data for a range of critical tasks performed by the Ops Room team, with a general focus on the role of ORO. A series of data collection trials is proposed that will match the developmental sequence of the data fusion and other technologies that form the basis of the COMDAT1 Technical Demonstration (TDP). This sequence will involve (i) an initial concentration on the integration of above water warfare (AWW) sensors and Link 11, followed by (ii) integration of surface aspects of AWW sensors, Link11 and underwater warfare (UWW) sensors, (iii) integration of all UWW sensors, and (iv) all of the former integrated with the wide area picture (WAP).

For each trial, a detailed description is provided of the logistics (personnel, software support and facilities) that will be required to prepare, conduct and analyse the trial. Trials will be scenario-based, with the scenario content largely focussed on the warfare area that the TDP is designed to support. The trial sequence will involve the initial collection of baseline performance data using existing technologies to serve as a subsequent comparison for the same tasks aided by COMDAT. Requirements for the test scenario are provided in detail, including the geographical, environmental, political and military contexts, the game entities and their dynamics and the sequence and number of events. It is recommended that these initial trials be conducted in the NCOT facility, which appears to provide all of the requirements for the first stage of evaluating the most immediately emerging TDP functionality, while minimising the logistical overhead to support the trial.

Details are provided of the processes required for the development of the scenario, its implementation in real time and the collection of reliable performance data using a broad spectrum of data capture methodologies.

The plan outlines a number of detailed measures of performance (MOPs) derived largely from a previous Cognitive Task Analysis of the ORO. The initial effort has been to develop MOPs that will focus on AWW and also for the generic situation awareness needs of the ORO in assessing threats and building the tactical picture.

The plan provides for details of the research design in terms of the number of data trials and test participant subjects that will be necessary to achieve the required sensitivity to detect effects of interest. In this respect, the Navy and the Scientific Authority are encouraged to think about the magnitude of effects that could be gained by COMDAT that would be of operational significance.

While the general strategy has been to mitigate risks by taking an incremental approach to T&E, certain risks have been identified. These include: (i) the ability of the NCOT facility to support team based scenarios and capture the required T&E data, (ii) the availability of an adequate number of OROs to act as test participants, (iii) the exact manner in which the TDP will be implemented, (iv) the possibility that operational realism may have to be traded against the requirement to capture reliable and valid data and (v) the generalisability of the findings to broader mission contexts.

THIS PAGE INTENTIONALLY BLANK

Abstract

This report provides a comprehensive Test and Evaluation (T&E) plan for evaluating the impact of future command decision aid technologies (COMDAT) on human and operational performance in the Halifax class operations room. The plan provides details of an incremental approach to T&E that will generate reliable and valid performance data for a range of critical tasks performed by the Ops Room team, with a general focus on the role of ORO. Elements of the plan include: the number and types of test trials, scenario development and implementation, requirements for personnel, logistics and facilities, details of the research design and a comprehensive list of measures of performance and the means by which data will be captured and analysed.

THIS PAGE INTENTIONALLY BLANK

Table of Contents

1. Introduction

This report represents a continuation of work aimed at improving decision support to the Operations Room (Ops Room) functions of the Navy Halifax class ships. The present work arises out of a contract from DCIEM to Human*systems* Inc. to develop a comprehensive Test and Evaluation (T&E) program to measure Ops Room functions and assess any future impact by MSDF technology.

This technical memorandum addresses specific Statement of Work items:

- 3.1.2 Develop Potential Measures of Performance (MOPs)
- 3.1.6 Create plan for piloting data collection
- 3.3.4 Create detailed evaluation report.

In this section the general approach and strategy to developing MOPs are described and issues concerning terminology and usage of the term "pictures" are addressed.

1.1. Developing the MOPs

The approach to developing the MOPs has been guided by a number of relevant sources that include:

- The ORO cognitive task and function flow analyses (references 1,2)
- A brief review of the Navy C2 literature concerning concepts of MOPs
- Previous analysis of methods for evaluating and measuring C2 system performance (reference 3)
- The COMDAT1 Technical Demonstrator (TDP) system design (references 4,5).

1.1.1 The ORO Cognitive Task Analysis

The foundations for the current work are based upon a cognitive task analysis (CTA) conducted with a representative cross section of serving Operations Room Officers (OROs) (reference 1). A scenario based walkthrough of a hypothetical mission, somewhat similar to Op Crater[1], was used to elicit the major operational goals that were required to achieve mission objectives. The information, situation awareness and communication requirements to achieve these goals were also identified. Initially goals were grouped into three categories: Coming on Watch, Foreground and Background goals. Subsequently, a validation of the initial findings was conducted on a different group of OROs (reference 2) As a result of this, Foreground and Background Goals were revised to become Threat-Related Goals and Ops Room Management goals. In addition, function flow analyses were created based upon the goal structure and SME input.

Using these analyses, the critical, core functions and tasks performed by the ORO to support generic mission goals were identified. These then served as the basis for focussing current efforts in determining appropriate MOPs that could be used to measure ORO/System performance in these areas.

[1] A scenario used by the Navy in Ops Room team training

1.1.2 Existing MOPs in the Literature

A brief literature review was conducted of military and particularly Navy MOPs that would be directly relevant to the current task. The focus of the review was on process-related measures rather than outcome measures of effectiveness (e.g. number of targets destroyed, number of casualties taken). In general, nothing specific was found to be applicable, other than recommendations to use accuracy/error and time on task measures, supplemented by subjective ratings where appropriate. Therefore, previous work on the measurement of C2 effectiveness was used to guide the selection of MOPs with a focus on measures of situation awareness, communication and decision making (reference 3).

1.1.3 Previous Analysis of Methods for Evaluating C2 Systems

A comprehensive review has been conducted on a variety of methods and measures that have been discussed or implemented in the search for appropriate ways to assess C2 performance in a broad spectrum of military contexts (reference 3). In general, the present approach follows the recommendations of this report by focussing largely on measures of performance that assess the underlying human-system processes that determine operational outcomes. The three process areas that are the focus of attention in the present analysis are: Situation Awareness, Communication and Decision Making. Specific methods and approaches for gathering performance data in these areas are provided in this report.

1.1.4 The COMDAT1 Technical Demonstrator (TDP) System Design

Another important element in focussing the scope of the *initial* set of MOPs has been the COMDAT1 TDP. At present, the TDP in many areas of MSDF[2] appears to be a concept in technical evolution, at least as far as we have been able to ascertain from available documentation and discussions with the Scientific Authority. Thus, there remains some uncertainty as to how and what functions will be implemented at the level of the ORO and other members of the Ops Room teams. Thus, while the focus of the CTA and the present mandate are the functions of the ORO, it would appear that in the initial stages the TDP effort may have more impact upon tasks performed by Ops Room teams in the different warfare areas.

Notwithstanding this uncertainty, the strategy that we have adopted in developing the MOPs is to match the ORO CTA goals with the COMDAT1 system concept as it stood at the time of writing. This allows the ORO functions to be identified that will most likely be impacted by MSDF concepts embodied in the initial TDP. From these functions we can then look at the specific tasks with a view to developing the appropriate MOPs.

The review of the available literature and technical memoranda and the information gathered from technical briefings indicate that, in the first instance, the TDP will focus on the integration of air sensor information at the track level, to aid tasks related to track detection, maintenance and possibly classification. As noted in previous discussions with the relevant Scientific Authorities, and based upon the CTA, normally the ORO has little direct involvement in the processing of information at this level or in the direct conduct of these tasks. Rather, the ORO's interest is managing processes performed by various members of the Ops Room team. The ORO will also under some circumstances share some of the air warfare duties with the SWC when the workload exceeds the capability of the SWC to cope. In the role of manager of the teams, the ORO has a

---

[2] We use the definition of MSDF as outlined in reference 4, multi-*source* data fusion (i e not just *sensor*)

major interest in the product of the various teams' activities in terms of building his tactical situation awareness, assessing and evaluating the tactical situation, planning actions, monitoring responses and co-ordinating the Ops Room surveillance function. Other aspects of the TDP, as outlined in the system concept (reference 5, figure 1) appear to have a more direct impact on some of the ORO core functions. These include the integration of the wide area picture, tactical picture, and the above and underwater pictures. This concept diagram also implies the integration of *non-sensor* sources of information. Exactly what technology is involved and how it will be implemented remains unclear at the moment, therefore we have made our best guess about the potential impact upon the ORO and have developed MOPs accordingly.

Further into the future, beyond COMDAT1, there remains the possibility that information fusion may impact tasks such as managing various aspects of the Ops Room capability (process refinement in the JDL model), which are core ORO functions. For now, we have put these issues on the back burner, given the immediate priorities of assessing the potential impact of MSDF in the COMDAT1 TDP.

Before proceeding with the description of the MOPs, we believe it would be useful at this point to clarify some terminology issues concerning the use of the word "pictures" to avoid subsequent potential confusion in interpretation.

1.2 Terminology Concerning "Pictures"

With respect to current Navy C2 thinking[3], the term Maritime Tactical Picture (which was used extensively in the CTA and reflected existing Navy terminology) appear to be have been replaced by the term Recognised Maritime Picture. The latter is part of a hierarchy of "pictures" as outlined below.

- Common Operational Picture (COP)
- Recognised Maritime Picture (RMP)
  - Maritime Surface Picture (MSP)*
  - Maritime Sub-surface Picture (MSubP)*
- Recognised Air Picture (RAP)
- Recognised Land Picture (RLP)

*Note: the above abbreviations represent labels of convenience for this report and do not necessarily reflect Navy terminology.

The *Common Operational Picture* is the composite of all operational information provided by each element (Maritime, Air, and Land)[4], fused and filtered as necessary to support the requirements of each specific user. The subordinate Recognised Pictures represent the operational information generated within each element, and specifically tailored for use by that element. For example, in addition to data on ships and submarines, a Maritime Commander's picture includes data on air forces (ASW helicopters, Maritime Patrol Aircraft, tactical fighters in combat air patrols over ships, hostile strike aircraft, missiles, etc). Data on ships is also of interest to an Air Commander operating near maritime areas. A Land Commander operating near coastlines will also be interested in data pertaining to naval and air forces in his area. There could also be other

[3] Based on references 6 and 7 and information provided by the Scientific Authority
[4] This includes the wide area picture (WAP), local area picture (LAP) and organic and non-organic sources

Recognised Pictures in addition to the above, and other sub-sets within each recognised picture. The underlying concept is that each picture represents the view of interest to a specific Command level; the COP represents the composite of all available data.

Formally, the RMP has been defined as "....a compilation of all source data relating to a specific ocean area, known at a given time and disseminated following evaluation and validation....". The term "recognised" means that the plot is an evaluated and validated interpretation of available information...". This can result in a recognised contact remaining "unknown" after the evaluation process.

Based upon the above, it would appear that the ORO will be dealing with pictures below the level of the COP, for the most part. The ORO's primary interest and responsibilities will be at the RMP level with a frequent requirement to interact with pictures below this level. Presumably, if TG duties included responsibilities for air warfare, then the ORO concerned would also have a primary involvement at the RAP level.

These definitions and concepts of the various categories of pictures will be subsequently used in this report as we identify the areas for specific MOPs.

2. ORO Functions Impacted by MSDF in COMDAT1

In this section we describe how we used the functions resulting from the CTA and CTA validation as a starting point for determining areas for which MOPs should be developed. From this initial set of functions, individual functions were either eliminated or augmented. Reasons for elimination included low relevance to MSDF improvements, low criticality or frequency and logistical considerations with respect to implementation for T&E. The list was augmented to include more detailed functions in specific areas that were to be the immediate focus of the COMDAT1 TDP.

2.1 Function Selection and Filtering

The first step in the process of determining MOPs was to identify the major ORO goals and functions from the CTA validation study that could be affected by MSDF technology, considered in its broadest form. These functions are shown in Table 1 and are taken directly from the function flow diagrams using the same reference numbers (reference 2 Annex A). Each function was then rated by the HSI team (which included one experienced ORO) in terms of operational criticality and frequency and the likely impact of MSDF technology. Then, in order to focus the MOP development effort, functions were eliminated from this list if they met any of the following criteria:

1. Low impact from MSDF technology (at least, as presently conceived in COMDAT)
2. Tasks done off line from the CCS
3. All tasks involving preparation and planning, which are considered to be out of scope by the Scientific Authority with the present focus on operational performance.

Functions eliminated by this process are shown with a grey background in Table 1.

Among the items eliminated by this process are a number of critical, **process management** functions that are core to the OROs overall operational focus. These have been temporarily set aside for now, for two reasons. First, we believe them to be generally beyond the *immediate* impact of the COMDAT1 TDP. Second, we have not seen any specific conceptualisations of the kind of COMDAT technology that could support ORO decision making in the conduct of these functions. The functions temporarily set aside are:

- Assess teams (2.4.1.4)
- Assess comms (2.4.1.5)
- Assess schedule (2.4.1.6)
- Optimise capability (2.4.1.7)
- Plan ship response (2.4.3.2)
- Implement response (2.4.3.3)
- Assess ship response (2.4.3.4)

If future TDPs incorporate some of the ongoing R&D concepts from DREV concerning situation analysis, impact assessment and process refinement, then the development of a wider range of MOPs for functions related to monitoring and managing processes and people will be pursued.

ORO Function/Task from CTA Goal List or Function Flow (ref.1)	MSDF IMPACT	CRITICALITY	FREQUENCY
1. Prepare for watch			
Review external environment - 1.1.1	LOW	MED	LOW
Update awareness - 1.1	LOW	HIGH	LOW
Review schedules - 1.1.2:	LOW	HIGH	LOW
Review capability - 1.1.3	MED	HIGH	LOW
Review ongoing Ops Room team tasks (2-5 mins) - 1.1.4	LOW	HIGH	LOW
Confirm authority - 1.1.5	LOW	HIGH	LOW
Review Mission Picture - 1.1.6	HIGH	HIGH	LOW
Comprehend briefing from outgoing ORO - 1.1.8	LOW	HIGH	LOW
Modify pre-plan/procedures for watch - 1.1.9	LOW	HIGH	LOW
Prepare workstation 1.2	LOW	HIGH	LOW
Focus OR team on issues for coming watch. (2 mins) - 1.3	LOW	MED	LOW
2. Conduct watch:			
Build /Maintain Global Picture 2 1 (see items below)			
Check Mission Picture- 2 1 1	HIGH	HIGH	HIGH
Relate (new info) to Ops Room picture - 2 1 2 (includes helo, weapons, CCS/SSD, organic sensors, comms)	HIGH	HIGH	HIGH
Relate (new info) to Ship Picture - 2 1 3	MED	MED-HIGH	MED
Relate (new info) to Task Group Picture - 2 1 4	MED	HIGH	MED
Relate (new info) to RAP 2 1 5 and RMP 2 1 6 (formerly referred to as MTP)	HIGH	HIGH	HIGH
Relate (new info) to environment picture - 2.1.7	ORO LOW Operators HIGH	HIGH	MED
Relate new info to ship schedule - 2.1 8	LOW	HIGH	MED
Prioritise mission needs -2.1 9	HIGH	HIGH	MED
Check own authority to act - 2.1.10	LOW	HIGH	MED
Plan watches ahead 2.2	LOW	MED	LOW
Manage Own Information Exchange - 2.3	LOW	HIGH	HIGH
2.4 MANAGE SHIP [(1)]			
Manage Ship Capability -2 4 1 (see items below)			
Assess sensors - 2 4 1 2	HIGH	HIGH	MED
Assess weapons - 2 4 1 3	HIGH	HIGH	MED
Assess teams - 2.4.1.4	LOW	HIGH	MED
Assess communications – 2.4.1.5	HIGH	HIGH	HIGH
Assess schedule - 2.4.1.6	LOW	HIGH	MED
Optimise capability - 2.4.1.7	HIGH	HIGH	MED
Manage ship surveillance - 2 4 2	HIGH	HIGH	HIGH
Manage ship response (see items below) - 2.4.3			
Ensure contact classified – 2 4 3 1	HIGH	HIGH	HIGH
Plan ship response - 2.4.3.2	MED	HIGH	HIGH
Implement response - 2.4.3.3.	MED	HIGH	HIGH
Assess ship response 2.4.3.4 (see below)			
Determine outcome for contact of interest 2.4.3 4.1	HIGH	HIGH	HIGH
Determine outcome for own ship 2.4 3.4.3	HIGH	HIGH	HIGH
Report- 2 4.3.5	MED	HIGH	HIGH

Table 1: Functions derived from the CTA and CTA Validation rated in terms of mission criticality and frequency and potential MSDF impact.

Notes:

(1) This analysis is based upon a threat-related mission

We now offer a few points of commentary on two of these decisions.

Optimise capability may have a high MSDF impact, however, the actual role MSDF may play in facilitating the performance of this function is not clear. It is possible that the impact of MSDF will be more in the form of assessing capability rather than assisting the process of rectifying any capability shortcomings. The assessment of capability can be thought of as having two parts: the assessment of sensors and the assessment of the deployment of appropriate personnel for the tasks at hand. There appears to be a clear role for lower MSDF levels for the former, but the latter would involve MSDF level 4 processes refinement, for which there appear to be no specific technological solutions on the immediate horizon. Thus, we have not developed specific MOPs for this function right now, but will use the MOPs associated with the function "assess sensors" to address the issue of assessing capability, where appropriate.

Determine outcome for COI. Clearly, this function has possible MSDF implications for the improved recognition of the changed status of hostile air tracks that might result from an engagement. As such it might be seen to have a high priority for MOP development. However, our reasons for setting it aside for now are largely practical, since it will demand a greater level of complexity in the running T&E trials, than the payoff may warrant. The additional Ops Room team and ORO functions that would be required to simulate the engagement of a threat presents a major additional overhead when considering the infrastructure and logistical requirements for the T&E scenario. Further, we believe that the specific changes in air tracks that arise from an engagement outcome can be simulated using the proposed threat events that are listed below.

Finally, the Scientific Authority may wish to evaluate priorities for the future development of MOPs in core areas of ORO functioning for some of those functions currently filtered out by the above process, even in the absence of conceptions of how technology could be of assistance. Several of these functions relate to the ORO's core role as the manager of processes and people and ultimately determine the overall *effectiveness* of the Ops Room.

2.2 Function Augmentation

Before proceeding further with refining the list, we saw a need to expand the depth of analysis for some the detailed Ops Room tasks that might be impacted by the more immediate implementation of COMDAT1 technology. This was considered necessary in view of the proposed focus of the application areas for the COMDAT1 TDP and some uncertainty as to whether the impact of the technology would fall upon the ORO or other members of the Ops Room teams. This analysis comprised a more detailed review of Ops Room processes with a Navy SME in order to gain a better understanding of where the COMDAT technology would impact. One result of this was the development of process flow diagrams that identified the major Ops Room tasks in the detect-to-resolve process that appears to be major area affected by the COMDAT1 TDP (see Annex A).

As a result of this process and insights derived from analysis of the JDL MSDF concept, it became evident that some revisions would be appropriate to the original CTA decomposition to accommodate this new information obtained. Specifically, the CTA and function flow analysis in the area of situation and threat assessment needed to be updated and enhanced.

In the CTA and subsequent function flow analysis, situation assessment (and the associated sub-functions relating to maintaining the air, surface and sub-surface pictures) were seen to be implied within the overall function of "*Build and maintain global picture*". In retrospect, it would appear that the verbs "build" and "maintain" overlook the need for "comprehension" and "understanding" in terms of tactical relevance that are the very reasons for building and maintaining the picture.

Given the importance of comprehension of the built picture to the assessment of the tactical situation, we propose that situation assessment should become an explicit focus for MOPs.

Situation assessment may be seen as a generic term that could be used across a variety of combat related or peacetime scenarios. However, in the context of typical training scenarios, situation assessment largely involves a narrower set of goals dealing with the evaluation of *threats*. We will therefore adopt this more restrictive usage as the basis for conceptualising MOPs, since the present focus of MSDF technology appears to be on improving information to the Ops Room team concerning contacts that are specifically threat related.

In terms of the existing function flow analysis, the "*assess situation*" function should probably be inserted as an explicit step after functions 2.1.2-2.1.8 (relating information to various pictures) and before 2.1.9 *Prioritise mission needs* and 2.1.10 *Check own authority to act*. Thus, in the tables and sections that follow, MOPs for threat assessment will be provided even though this function does not map directly onto the organisation of the functions provided by the CTA or function flow analysis.

A second area that we believe to be of importance that was captured in the original CTA, but not reflected specifically in the function flow diagrams, concerns the issue of *attention switching by the ORO*. This was identified as a critical ORO function with respect to both the evaluation of the tactical picture across all warfare domains as well as in managing and monitoring Ops Room processes and personnel. Therefore, we have included attention switching as a potential area for MOP development.

A third area of potential focus for MOPs resulting from a review of the COMDAT1 TDP system concept concerns the integration of wide area picture (WAP) information. This was not explicitly differentiated from the Air, Surface or Sub-surface pictures in the CTA or subsequent validation. However, given that the TDP will be designed to facilitate the integration of the wide and local area pictures, no matter the domain, it would be prudent to consider this as an area for performance assessment and hence MOP development.

To summarise, the above list of functions from the CTA validation has been augmented in the following three areas:

- Threat assessment
- Attention switching
- Integration of information from WAP

We now return to the issue of MOP development.

2.3 Priorities for MOP Development

Table 2 shows the list of functions resulting from the initial filtering and augmentation process outlined in 2.1 and 2.2. In order to establish priorities for the initial development of MOPs for these functions, it has been necessary to make our best guess as to what MSDF technology is likely to emerge in the immediate future. In the absence of any other concrete conceptualisation, we have used the DREA System Concept (references 4,5) for guidance, which outlines the following areas that will be the focus of COMDAT1. The ordering below is based upon information that has been provided by the Scientific Authority on the planned development and implementation sequence for the TDP.

1. AWW tactical picture: implications for RAP, MSP
2. Integration of AWW/WAP: implications for RAP, MSP, RMP

3. Integration of AWW, UWW and WAP: implications for RAP, MSP, MSubP, RMP.

Before relating these areas of focus of the TDP to the Ops Room Functions, some elements of uncertainty surrounding the above should be noted. With respect to step 1, the documentation that has been made available to us, and information communicated from the Scientific Authority concerning the trials that have been conducted, suggest that the TDP emphasis has been on the fusion of *all air sensor and Link 11 data*. This would be of primary relevance to AWW. We have found no mention of the specific Navy systems that would form the basis for fusion of data, or integration of information that would address the other component of AWW, namely the MSP. A further consideration that requires clarification concerns the integration of the UWW sensor information in step 3. There are at least two possibilities. First, by analogy with the integration of air sensor data, the UWW sensor data (CANTASS, HMS, SPS) could itself be integrated to provide the best available UWW picture. Or, second, the existing streams of already processed information from these three systems could simply be provided at the track level and displayed upon a common integrated display that would also include air and surface tracks. Clarification of these areas will be necessary at some point in time to ensure we fully understand which Ops Room functions and personnel will be impacted by the various possible configurations of data and information elements.

Notwithstanding this ambiguity and uncertainty, our next step has been to map the three TDP areas against each of the remaining subset of ORO functions in terms of where the TDP is likely to have impact. The outcome of this is shown in column three of Table 2. Where we could not see an obvious impact of the COMDAT1 technology, we have indicated not applicable (NA). We have also indicated the appropriate domain of measurement[5] in column four, where SA= situation awareness (levels 1,2,3[6]), COM= communication (that is all aspects of information exchange, not simply those aspects using only audio comm nets) and DM= decision making.

The final column of Table 2 provides a recommendation of priorities for developing the MOPs. These recommendations are based upon a consideration of four factors. First, the importance and criticality of the function and the degree of MSDF impact. Second, our current perception of the timeline for the TDP development - functions that are likely to be impacted earlier are assigned higher priority. Third, our intuition about the difficulty in implementation of some of the associated, core technologies and the clarity of the concepts. Fourth, the logistical difficulty in collecting MOP data - functions that will require complex scenarios across several warfare areas and involving several Ops Room teams are assigned lower priority.

It should be noted that in this table we have changed the function 2.1.6 *Relate to MTP to Relate to RMP* to be consistent with more current Navy terminology. We have also further decomposed this into the constituent elements of "Relate to... Maritime Surface Picture (MSP)",.... Maritime Sub-Surface Picture (MSubP) in order to provide a better mapping onto the separate functional areas of the COMDAT1 TDP. We have also included the WAP in the consideration for developing MOPs, even though this was not identified as such in the CTA, since this is clearly a focus of the MSDF technology and can be considered as contributing to the RMP.

5 Based upon reference 3

6 SA1=detection of new information, SA2=integration and comprehension of information, SA3=projection of future state

T&E Task#	ORO Function/Task from CTA Validation Function Flow	COMDAT1 Reference (as above)	Focus of MOP from Framework for Evaluation	Priority
	2.1 BUILD MAINTAIN GLOBAL PICTURE			
1	Check Mission picture - 2.1.1	1,2,3	SA1, COM	1
2	Relate new info to Ops Room picture - 2.1.2	1,2,3	SA2,3 COM DM	1
NA	Relate new info to Ship Picture - 2.1.3	NA	-	-
NA	Relate new info to Task Group Picture - 2.1.4	NA	-	-
3	Relate new info to Recognised Maritime Picture (formerly MTP)	1,2,3	SA2,3	1
4	Relate to Recognised Air Picture -2.1.5	1		1
5	Relate to Maritime Surface Picture and Sub-surface Picture -2.1.6	1,3		1
6	Relate to Wide Area Picture*	2,3		3
7	Assess threats*	1,2,3	Implicit in SA2	1
8	Switch attention between different pictures**	2,3	SA2,3	3
	2.4 MANAGE SHIP			
9	Assess sensors - 2 4 1.2	1,2,3	SA1 DM	1
NA	Assess weapons - 2.4 1.3	NA	-	-
10	Manage ship surveillance - 2.4.2	1,2,3	SA1,2,3, DM, COM	2
4***	Ensure contact classified – 2.4.3.1	1,2,3	SA1,2,,3 DM, COM	1

Table 2: Mapping of COMDAT1 Technology onto ORO functions, evaluation focus and priorities for MOP development.

Notes on Table

* No specific counterpart in CTA function analysis
** Taken from CTA, this not explicitly identified as a unique function in the function flow analysis
*** To be measured as part of task 4

2.3.1. Rationale for Final Selection of Tasks to be Measured

A number of the functions in Table 2 have been assigned lower priority for measuring, despite having high criticality and medium or high MSDF impact This section provides the rationale for those assignments.

Switch attention between different pictures is an ORO process identified in the CTA that falls within the "Build and Maintain Global Picture" function but could not identified as a unique sub-function in its own right in the function flow diagrams. While we believe this process to be one of the most critical components for effective ORO performance, the rating of 3 has been assigned because of logistical considerations in its measurement. In order to elicit the behaviour of switching between warfare domains, a greater complexity of scenario will be required to include events of interest across all warfare domains. We believe that we will be in a better position to formulate MOPs for this function, once some experience has been gained in scenario development, data collection and proof of concept testing for MOPs using less complex scenarios. Lessons learned should then enable us to rapidly formulate the most appropriate MOPs for this function.

Relate information relevant to the WAP would seem to be directly impacted by the intention in COMDAT1 to integrate the WAP and LAP pictures, by making available to the ORO information from the GCCS/ MCOIN systems. However, we have not developed specific MOPs for this function because at present the concept of how this is to be realised seems somewhat unclear. For example, the GCCS currently provides *information* (that is processed tracks) rather than *data* (radar returns). Hence, it is not clear how such distal information can be "fused" with local radar data - since they are addressing essentially different subsets of contacts at different ranges. Our current belief is that the proposed approach in the TDP for the provision of the WAP to the ORO appears to be more in the format of *information aggregation* rather than data fusion. One design approach that is being considered to implement this concept is to provide dual displays - with the lower providing information on the LAP and the upper the WAP. That is, information at the track level for the wide and local pictures will be co-located. This approach is somewhat different from concepts that involve *fusion* or *integratation* at the semantic level. This design solution would improve upon the current situation whereby the LAP and WAP are displayed on monitors that are in different places in the Ops Room. However, it remains to be seen whether this relocation of displayed data results in providing the ORO with a seamless view of the local and long-range tactical pictures, as described by the "fused scene" concept in reference 6. Notwithstanding this uncertainty concerning the level at which data or information are to be integrated and the exact design solution that will be adopted, it will probably be the case that MOPs developed and tested for somewhat similar functions relating to the detection, integration and interpretation of new information for the RAP or MSP or MSubP may be readily modified to assess integration of the LAP and WAP.

Manage ship surveillance is also rated HIGH for MSDF impact, where we believe that the contribution of the technology is likely to be in the area of providing better information to the ORO concerning the performance of sensors and integration of information from the WAP and across warfare domains. Many of these capabilities can be assessed by MOPs outlined elsewhere for related functions. The other aspect of managing surveillance concerns the deployment and monitoring of personnel performing surveillance functions and the development of surveillance plans in light of changing circumstances. For the former, a preliminary set of MOPs is suggested relating to the ORO's ability to observe when there are errors or problems in the standard detect-to-resolve process (in the specific case of air threats).

2.4 Recommended MOPs

The initial focus of MOP development is on the functions to which we have assigned a priority rating of 1 or 2 in Table 2. In specifying the MOPs in the table that follows, some of the above functions have been grouped together because of overlapping MOPs. For example relating new information to the RMP is highly related to the function of building awareness across domains and forming an integrated tactical picture. In this initial set of MOPs, greater attention has been placed upon the function "Relating new information to the RAP" since integration of air sensor data this will be the initial focus of the COMDAT1 TDP. This function has been further analysed to a greater level of detail than the original CTA in order to provide specific, process-oriented MOPs that correspond to the major tasks performed by the Ops Room Team. This analysis was performed using a former Navy ORO who was part of the HSI team.

Note that no specific measures have been presently identified for functions relating new information to the Maritime Surface and Sub-surface pictures. We believe that the experienced gained in first measuring performance in the air domain and assessing the various possible measures will readily prepare us for subsequent T&E phases involving other warfare areas.

DETAILED MEASURES	COMMENTS
T&E TASK 1 Check mission picture 2.1.1	
1a Accuracy in detecting relevant info within incoming message stream	Manipulate info content relevant for circumstances, some immediate, some requiring later action. Vary workload Vary operational circumstances to influence salience of info Vary message source, TG, Ops Room, ship, GCCS Need SME judgement for ground truth
1b Accuracy in ignoring irrelevant info	Need SME judgement for ground truth
1c Total time spent in comms dealing with incoming info	Gross measure
1d Time to detect high salience message	For text – what specific event would be the marker for the start time Assume for voice message this is not a problem.
1e Accuracy in requesting additional info (i e number of messages that required follow up that resulted in ORO request for more info)	Ensure that some messages require a follow-up enquiry. May be difficult to assess unless there is a large message pool.
1f Accuracy in ignoring lower priority messages	Ensure that messages vary in priority May be difficult to assess unless there is a large message pool. Need SME judgement for ground truth
T&E TASK 2. Relate new info Ops Room picture 2.1.2	
2a Accuracy in directing (communicates) info/direction resulting from message	Could be voice or text action?
2b Accuracy re message content in briefing appropriate station	
2c Accuracy in comprehension of impact on pre-plans/response options/tactical situation	Need SME judgement for ground truth
2d Accuracy in recognition of impact on Ops Room capability	Need SME judgement for ground truth
2e Total time spent in relaying info re Ops Room status	
T&E TASK 3 Relate (new info) to RMP Includes aspects of build awareness across domains (RAP, MSP, MsubP). Same measures will apply in principle to *6 Relate to WAP*	
3a Accuracy of salient info within RMP *prior* to new message	Need ongoing regular probes of OROs knowledge of salient aspects of MTP.
3b Accuracy of salient info within RMP *after* new message	
3c Time for ORO to complete understanding of new info regarding RMP	T&E probe could be in the form of a CO standing order for a SITREP or pressing a function key when he believes new info changes tactical pic.
3d Accuracy in taking appropriate action as a result of updated RMP	Could provide info that requires specific action to be taken (e g change course, change direction to teams, advise CO/TG, review pre-plans etc. Hence we need to specify a wide range of salient information dimensions that will be manipulated

DETAILED MEASURES	COMMENTS
3e Total time spent on SSD relating new info to RMP	Need to define a trigger event and "end" event - could require a specific T&E probe etc.
3f Accuracy in assessing the integrated tactical picture	Freeze probes or SITREPS Includes knowledge of threats and contacts of interest in relation to ship or TG.
T&E TASK 4 Relate new info to RAP 2.1.5. (as an example): Ensure contact classified 2.4.3.1 Note Same approach/measures to be used in principle for MSP and MsubP	
SUB TASK 4.1 Identify friendly aircraft Friendly / neutral a/c will provide the majority of items in the air contact stream. There will be two conditions when no hostile a/c is being dealt with, and when one is being dealt with. Distraction / attention management factor will be quite different i e more likely to miss / short cut identification procedures when focussing on a hostile	
4 1a Accuracy in identifying friendly aircraft (by type/mission)	
4.1b Mean time spent in locating friendly aircraft	Must be expressed in relation to aircraft load factor
4 1c Total number of queries or total time spent in querying team for additional info	
SUB TASK 4.2 Identify hostile/suspect aircraft Hostile / suspect a/c will be occasional insertions in main air contact data stream.	
4 2a Accuracy in identifying hostile / suspect aircraft (by aircraft type)	
4 2b Mean time spent in locating hostile/suspect aircraft	
4 2c Total number of queries or total time spent in querying team for additional info	
SUB TASK 4.3 Identify neutral aircraft	
4 3a Accuracy in identifying neutral aircraft	
4 3b Total time spent in locating neutral aircraft	
4 3c Total number of queries or total time spent in querying team for additional info	
SUB TASK 4.4 Identify NU tracks (non-updated)	
4 4a Accuracy in identifying NU tracks	
4 4b Total time spent in locating NU tracks	
4 4c Total number of queries or total time spent in querying team for additional info	
SUB TASK 4.5 Identify tracks reported by ownship or, conversely, by other participating units (PUs)	
4 5a Accuracy identifying air tracks being reported by ownship	
4 5b Total time spent locating air tracks being reported by ownship	
4 5c Total number of queries or total time spent in querying team for additional info	
SUB TASK 4.6 Identify own force engagement status This measure will only be used if scenario requires engagement	
4 6a Accuracy identifying air tracks being engaged by ownship or ships in TG	Or other friendly a/c??

DETAILED MEASURES	COMMENTS
4.6b Total time spent in locating air tracks being engaged	
4.6c Total number of queries or total time spent in querying team for additional info	
SUB TASK 4.7 React to threat track symbology LINKed by consort not on video	
4.7a Time to recognise symbology not on video	
4 7b Time to order remedial action	
SUB TASK 4.8 React to LINK not gridlocked	
4.8a Time to recognise LINK not gridlocked	
4 8b Time to order remedial action	
T&E TASK 7 Assess threats -generic	
SUB TASK 7.1 Detect changes in tactical situation	
7 1a Accuracy in ID new threats	Probe data to be inserted by T&E team
7.1b Accuracy in ID changes in threat status	
7 1c Time to detect new threat	
7.1d Time to detect change in threat status	
7.1e Time to provide SITREP	Request for SITREP provided by T&E team Can be done as freeze probe or as an ongoing request for info
7 1f Accuracy in SITREP contents	Need SME judgement for ground truth
SUB TASK 7.2 Determine threat priorities across domains Will need to vary circumstances to ensure that some threat changes occur in a particular domain while the focus is in that domain, at other times when focus is in another domain	
7 2a Accuracy in ranking threat priorities across domains	
7 2b Time to assess threat priorities across domains	
SUB TASK 7.3 Assess threats- (air warfare specific) Identify air track threat priority Similar MOPs should apply in principle to other warfare areas	
7 3a Accuracy in identifying threat priorities	Accuracy in terms of time to intercept or aircraft weapon lethality SME to provide ground truth
7 3b Total time to identify three highest priority threats	
7.3c Time to annotate CCS with CPA	Determine closest point of approach (CPA) for each threat
7 3d Accuracy in determining CPA	RMS lat/long measure
7 3e Accuracy in determining lethality	Determine 'lethality' for each threat SME to provide ground truth
7.3f Time to determine lethality	May need specific ORO action to indicate task completed
7 3g Time to create a SITREP	Appreciation of overall air tactical situation
7 3h Accuracy and completeness of SITREP	
SUB TASK 7.4 Analyse history profile[7] of hostile aircraft	

[7] By history profile we mean the patterns of trajectory shown over time by a particular contact that is potentially hostile, these patterns include changes in altitude, speed and direction, communication trends, EW emissions etc

DETAILED MEASURES	COMMENTS
7 4a Accuracy in analysing attack history of hostile aircraft	Number of fighter/bomber (FBA) runs, or missiles fired
7 4b Time to analyse attack history of hostile aircraft	
7 4c Accuracy in assessing number of weapons remaining on hostile aircraft	Assume a given quantity at game start as assessed by intelligence sources
7.4d Total number of queries or total time spent in querying team for additional info	
7 4e Accuracy in assessing attack tactics used by hostile aircraft (individual contacts)	Firing range, altitude profiles, number of weapons used
T&E TASK 9 Assess sensors Sensor information can become degraded through equipment malfunctions or restrictions. Hence, will need to *ensure that sensor becomes degraded during scenario*	
9a Accuracy assessing that sensors are less than optimum	
9b Time to appreciate that sensor is performing less than optimum	Need to log time at when sensor becomes degraded and ORO takes action by looking for relevant ORO comm.
9c Accuracy in identifying current sensor range predictions	
9d Total time in identifying current sensor range predictions	
T&E TASK 10 Manage ship surveillance	
10a Accuracy in recognising the problems in detect to resolve process	Need to ensure errors of different types presented (e.g. amplification, inappropriate work focus, communication errors)
10b Time to recognise problems in detect-to-resolve process	
10c Appropriateness of remedial action to correct errors	SME evaluation

Table 3: Details of MOPs and comments on their implementation

3. Test and Evaluation Plan: Overall Strategy

3.1 General Strategy

A number of factors influence the general strategy for developing the test plan. These include:

- An initial focus of the COMDAT1 TDP on AWW
- Current uncertainty concerning the way the TDP will evolve to impact upon WAP information integration, and surface/sub-surface integration
- The need to acquire experience in developing scenarios for the specific test environments contemplated (NCOT, ORTT) and running T&E scenarios in the specific test environments. (Especially for NCOT which has not been used for such complex scenarios previously).
- Uncertainties over the specific logistical requirements for human-centered[8] testing, until pilot trials have been conducted

All of which suggest that an incremental approach be adopted for testing both in terms of logistical scope and complexity of the domain to be studied, in order to minimise and manage risk. We propose that the initial area of focus start with the RAP, in later phases we will look at the MSP, MSubP and the RMP. In each case we commence with simple scenarios to allow proof of concept methodologies and gain experience, and then progress to more complex scenarios in terms of events and players, to full Ops Room team simulations.

3.2 Choice of Approaches

There are two basic goals for all human-centred testing: first to collect robust and reliable baseline data for core and critical tasks, and, second to collect data for those same tasks as influenced by the TDP. There appear to be two possible approaches to data collection for the above purpose. First, to set up a T&E environment using one of the Navy simulation facilities and to conduct a series of trials specifically for the purpose of collecting MOP data. This report focuses on this approach and outlines the necessary detail to conduct the trials. A second approach would be to use existing data created by the Navy during training or work-ups. Our review of potential test environments showed that the ORTT facility is capable of running full operational scenarios involving the whole Ops Room team and recording in a high level of detail the actions, screen contents, communications of the team under study. We believe that the *records of such training sessions may prove to be a useful source of baseline information*, from which MOPs for a wide range of tasks could be extracted after the fact.

For example, by watching the RT1 screen during playback and following the AAW team communications it would be possible to time the detect to resolve cycle for a number of contacts of interest. Further, we would be able to monitor the ORO's interactions and interventions to ensure that processes are being conducted appropriately. We could also trace the ORO's actions in switching between different representations of the tactical environment on the CCS. Some of the measures extracted could be timed-based, others might be SME ratings or assessments of the ORO's actions and communications. Gross measures of communication could also be readily

[8] The COMDAT1 SOR refers to this as human-in-the-loop testing (HIL)

obtained as well as communication patterns between team members associated with specific scenario events. Our initial evaluation of the ORTT suggested that this information can be obtained with a sufficient degree of precision and reliability to meet the needs of T&E.

It will be recalled that the ORTT contains a group debrief facility with 3 large projections screens and loudspeakers for audio playback. In addition, data recorded during a training session may be replayed in the Training Control Facility. The system provides a debrief mode which allows replay of a simulation at one of four (0.5, 1, 2 or 5) play speeds for the debrief session of data that have been recorded. A Debrief Control is a tool available to the training staff to debrief trainees on completion of training sessions. It uses game data saved by the Game Monitoring functions. Three monitors, an audio system and a screen projector are available to debrief the trainees. Information including hardware panels, CCS data, synthetic environment, Link-11, CCS tactical pictures, trainee consoles and audio recordings can be presented. Editing tools are available to assemble and compile an audio/video presentation from the recorded data. There is a capability to playback the data at slower or faster than normal speeds and to jump to specific points of interest that have been tagged either prior to, or during, the scenario execution.

Given the potential usefulness of the ORTT training records, *we recommend that a more in-depth evaluation be conducted to determine if useful MOP data could be extracted.* To achieve this, HSI® staff would conduct a site visit to the ORTT and review as a first step the types of scenarios and associated records that are available. The second step would be to select an appropriate sample playback and conduct a proof of concept analysis in terms of extracting potential MOP data Access to such scenario records could be scheduled at off-peak hours[9] or during the eight hours per day set aside for software development. This site visit could be integrated with a review of an actual exercise in progress in the ORTT, the justification for which is outlined in section 5.7.4.

Since the potential usage of the ORTT to collect MOP data remains somewhat speculative at the present time, the balance of the section on T&E strategy will focus on providing details of a test program to collect real-time MOP data in a series of dedicated trials in NCOT.

3.3 Sequence of Major Test Trials

The above considerations suggest an approach that follows the following sequence of major trials:

1. Collect baseline performance on ORO tasks relating to the RAP. Measures 4,7.3,7.4,9, 10.
2. Repeat 1, as influenced by the TDP.
3. Collect baseline performance on ORO tasks relating to the generic detection of new information (i.e. not just air surveillance) and on ORO tasks relating to the MSP and MSubP. Measures 1,2,7,9 (adapted for other warfare areas).
4. Collect baseline performance on ORO tasks relating to the integrated RMP and WAP Measures 3. 7.1,7.2.
5. Repeat 3, as influenced by the TDP.
6. Repeat 4, as influenced by the TDP.

This sequence should be considered as having some flexibility depending upon the actual progress made in developing the TDP. For example, trials 3, 4 could be brought forward, if the technology were not in place to perform trial 2.

---

[9] Note that the ORTT can only be used in either game playing mode or playback mode at any one time Off peak access would therefore minimise conflict with Navy training requirements

Each of the test trials will have a structure that involves at least the following elements:

- Logistical requirements
- Scenario development
- Detailed test plan
- Proof of concept
- Pilot study
- Main study
- Data analysis
- Conclusions and lessons learned

In the following sections, we outline some of these basic requirements for each of the above in as much detail as is currently feasible, given the current state of knowledge and experience.

THIS PAGE INTENTIONALLY BLANK

4. Logistical Requirements for Collecting MOPs

This section outlines the T&E infrastructure that will need to be in place in order to support the first test trial. One major uncertainty concerns whether all events can be pre-programmed into a scenario to run on a predetermined schedule, or whether live players of certain roles and "drivers" of data will be required. For example, in the case of ambiguous air tracks, it is not known in NCOT or the ORTT simulators, whether the kinds of operational behaviour of sensor systems found during actual sea conditions can be replicated. If not, live game players will be required to drive tracks in a particular way in real time in order to simulate real events[10]. Such uncertainties have major implications on the logistical overhead for the running of a scenario and the implementation of the scenario elements into the appropriate software. Such requirements will become better known after the first proof of concept trial.

In general, each trial will comprise following T&E phases.

4.1 Preparing the Trial

Naval liaison: This will be required to book the facility, arrange for T&E staff visits and access, arrange for test participants with appropriate naval SME background.

Scenario development and encoding: T&E personnel will require access to existing scenarios and related OPGENs and OPTASKs. T&E staff will need to be trained in scenario development and modification. If current OPTASKs are not available, Navy personnel will be required to assist in the development of this information. Software specialists will be needed to ensure that the scenario is pre-tested and functions appropriately. Access to suitable workstations may be required for T&E staff to proof and test scenario components. Naval SMEs will be required to provide some expertise on the events to be simulated (wherever possible this will come from personnel within the T&E team).

Software development and coding: Some modifications to the existing simulation software may be required to support T&E requirements. These might include the ability inject flags into the scenario database to signal start of key events, to provide a capability for injecting T&E information probes into the scenario in real time, to capture and log the time of participant actions at a workstation, to freeze the scenario and restart. The responsibility for defining the requirements lies with the T&E team, the responsibility for any software coding to incorporate the requirements will be with the software developers for a particular facility, for which suitable budgeting and contracting provisions will be needed.

4.2 Running the Trial

Information sources: these should represent all aspects of a normal Ops Room where the ORO can be expected to gain information. Sources of text (CCS or paper) and audio messages include the

[10] As a result of a recent evaluation of the NCOT facility (see Annex B), we know that an RTI will be required to enter radar tracks but that certain aspects of radar returns from contacts will need to be simulated using real-time control of game entities by the T&E team

ongoing Ops Room team, the TG, other areas of the ship, the GCCS and stateboards and any other of the relevant communication nets.

Ops Room personnel-real or simulated: these represent "live" players with whom the ORO would normally communicate and interact. For most of the tasks anticipated to be impacted by the TDP, we would expect the major players to be the SWC, ASWC, CO and ORS. Information flow from other members of the Ops Room including the front row team and sonar systems normally flows to the ORO through the warfare directors (WDs).[11] Based upon a recent, in-depth evaluation of the NCOT environment (Annex B), we believe that experienced Navy personnel will need to play the roles of the RT1 and SWC and/or ASWC when the trial involves other than air warfare. An experienced RT1 will be required to process the basic sensor data that comprise the radar picture that is used by the ORO and SWC.

In general, the dynamic information to be provided to the team that is beyond that contained within the scenario pre-scripted events and data will come from one of two sources, T&E personnel who follow a precise script and a Navy SME (part of the HSI® team). The latter will play many roles by providing all of the technical communication to the ORO (e.g. while acting as CO), by manipulating the tactical picture in real time as circumstances warrant, and by generally providing all specialised, knowledge-based information that cannot be pre-planned but is required for a realistic scenario. Another role for this individual will be to observe and make notes, for later analysis, on the actions of the ORO at certain times. The individual who plays this role must have an intimate knowledge of ORO functions, Ops Room procedures and expected performance for the particular scenario. For convenience, and given the omnipotent, all-knowing insight required of this role, we refer to them subsequently as the Gaming Operational Director (GOD).

Workstations: these are the physical simulations of actual Ops Room workstations that will need to be in place, in order to allow the associated Ops Room team member to perform their normal tasks. For the most part, we anticipate that most of the human-centred testing for the COMDAT1 TDP can be accomplished with workstations for the ORO, SWC and ASWC plus additional workstations for team members such as the RT1 or SCS, depending upon the domain focus of any trial. An additional workstation will be required for T&E personnel to monitor events on any of the subject workstations. This dedicated T&E workstation should have the capability to display the following information: CCS Data, CCS Tactical Picture, Link-11, Synthetic Environment from the perspective of any member of the operational team (usually the ORO). In addition, from 1-3 workstations will be required for GOD and other T&E personnel in order to be able to inject dynamic, real-time data and messages during the course of the scenario.

Environment: Necessary parameters of the operating environment will need to be simulated where appropriate. This includes selected elements from the natural, tactical, EW and acoustic environments. Normally this capability is inherent in the simulation facility to be used.

Data to drive the simulation: This will either come from pre-scripted events or will be provided in real-time by game players who act as data sources or drive game entities following pre-planned scripts for the most part. Some of the data will be provided by Navy SMEs playing the roles of the SWC, ASWC and subordinate members of these teams, as the scenario context demands.

Simulation support: this includes all personnel required to support the trial (other than role players) and will include network technical staff, software staff and naval liaison staff.

---

[11] In later versions of COMDAT, when information related to process regulation and monitoring is developed, it may be necessary to have all Ops Room personnel represented This is because a major role of the ORO in his management capacity is process refinement

4.3 Analysing and Reporting the Trial

Data translation for analysis: If the output data from the trial is not available in a format that permits analysis in standard commercial database and spreadsheet applications, then support personnel will be required to provide this transformation capability.

Data interpretation and analysis: This will be performed by the T&E team supported by Naval SMEs in situations where qualitative evaluations of the captured data must be performed. Facilities that may need to be provided for this activity include workstations for the replay of data and/or scenarios and equipment for the replay of audio or video logs.

Reporting the trial: The T&E team will have the responsibility for providing a written report to the scientific authority and to provide an oral briefing of the results. It is expected that this would be performed after completion of each of the data capture phases outlined in section 3.1 above.

The following table summarises the major logistical requirements for the four ORO functional areas to be evaluated.

	Detect Incoming Information	Build/Maintain RAP	Build/Maintain MSP,MSubP	Build/Maintain integrated RMP/WAP
Information Sources				
Text messages	X	X	X	X
Link 11 (CCS symbology)	X	X	X	X
External net TG	X	X	X	X
Internal net Ops Room· General (C&C)	X	X	X	X
Internal net team (AWW)	X	X	X	X
GCCS (symbology/map/text)	X		X	
Stateboards	?			
Ops Room Team Players				
SWC	X	X		X
ASWC	X		X	X
ORS	X			X
EWS				
CO	X	X	X	X
ORO	X	X	X	X
RT1	X	X		X
SCS	X		X	X
TS	X	X	X	X

	Detect incoming information	Build/Maintain RAP	Build/Maintain MSP,MSubP	Build/Maintain integrated RMP/WAP
RT2			X	X
Workstations Real/Sim				
SWC	Real	Real		Real
ASWC	Real		Real	Real
ORS	NA	NA	NA	
EWS	SIM	SIM	SIM	SIM
CO	SIM	SIM	SIM	SIM
ORO	REAL	REAL	REAL	REAL
RT1	Real	Real		SIM
SCS	SIM	SIM	SIM	SIM
RT2	Real		Real	SIM
TS	Real			SIM
SAC	SIM	SIM		SIM
HMS				SIM
Source of data for events				
Text messages	Pre-scripted/T&E team	Pre-scripted/T&E team	Pre-scripted/T&E team	TBD
Link 11	Pre-scripted/T&E team	Pre-scripted/T&E team	Pre-scripted/T&E team	TBD
Internal net Ops Room	Pre-scripted/T&E team	Pre-scripted/T&E team	Pre-scripted/T&E team	TBD
Internal net ship	Pre-scripted/T&E team	Pre-scripted/T&E team	Pre-scripted/T&E team	TBD
GCCS	Pre-scripted/T&E team	Pre-scripted/T&E team	Pre-scripted/T&E team	TBD
Stateboards	Pre-scripted/T&E team	Pre-scripted/T&E team	Pre-scripted/T&E team	TBD
Sensor data - air	Pre-scripted/live player	Pre-scripted/live player	Pre-scripted/live player	TBD
Sensor data - sub-surface	Pre-scripted	NA	TBD	

Table 4: Summary of logistical requirements

THIS PAGE INTENTIONALLY BLANK

5. Detailed Approach to the Evaluation

An incremental approach has been suggested above for the conduct of test and evaluation and the collection of MOP data. This will ensure that risks are managed and resources applied wisely and minimally in the initial stages where feasibility is being confirmed. As more experience is gained in creating the necessary T&E environment and the collection of data, the scope of scenarios and range of MOPs may be expanded. The initial focus of the evaluation will be on measures relating to building and maintaining the RAP, since this is the function most immediately impacted by the COMDAT1 TDP.

Based upon our initial assessment of data collection environments, and a recent follow-up visit to NCOT to determine specific capabilities and limitations (see Annex B), *we recommend that the initial proof of concept and pilot data collection be conducted in the NCOT facility.* This environment appears to have the necessary capability for simulating air warfare involving the RT1, SWC and ORO as interacting players.

The first trial will involve the collection of MOP data for the function **"Build and maintain the RAP"** and will proceed in the following phases.

1. Generate software requirements for implementing proof of concept testing of environment and MOPs. Generate software.

This activity is expected to involve primarily HSI® staff and will result in the production of a requirements list to support T&E test trials for the software developers of NCOT. Any software that will be required to be developed to support the test trial (i.e. is outside of the current "build" capabilities of NCOT) will *need to be funded and contracted separately with McDonald Dettwiler.*

2. Generate scenario; encode scenario events in software.

By scenario we mean the sequence of events and associated contextual information that will form the basis for guiding the actions of Ops Room personnel.

HSI® staff will review existing scenarios that have already been created for training or other purposes to assess their suitability for T&E purposes with a view to modifying them as required. In the event that nothing suitable is available, then scenarios will be built from the ground up. Our preliminary assessment is that there is nothing currently available that has been pre-programmed for the NCOT environment to meet T&E needs, but a library of game entities is in place that can serve as the building blocks for generating scenario events.

In order to develop and test the scenario events, access to an NCOT workstation will be required for T&E staff. In order to reduce the travel overhead, inconvenience and other associated costs with doing this work at the NCOT facility, *we recommend that HSI® staff be given access to an NCOT workstation more locally.* This could be achieved by making an NCOT Unix workstation available at HSI® offices or at DCIEM. Whatever the location, we strongly recommend that such a workstation be acquired, since the short term cost of acquisition will be more than offset by future costs associated with travel to Halifax. Further, the provision of such a workstation locally would better support the continuing needs of the development of MOPs related to COMDAT and would allow the local testing of concepts and scenario events in an efficient and cost effective manner. Another major advantage would be in providing a local capability to playback and analyse data from T&E trials, again without the overhead of travel to Halifax.

Notwithstanding the location of this workstation, the encoding of the scenario events into software and their initial testing will require the support of Macdonald Dettwiler staff, and contractual arrangements will need to be made to formalise and fund this process.

3. Initial proof-of-concept assessment and familiarisation with the selected test environment (NCOT in the first instance)

This process is designed to provide an early check on whether the scenario will run as required; whether the real time probes can be injected; whether events can be captured, timed and logged; to work out logistics for personnel who will be driving data and playing roles and to ensure that stored data are amenable to analysis. It will also be used to evaluate whether the appropriate level of realism can be achieved for simulated events such as ambiguous radar data and loss of radar data. Major tasks are the preparation of the scenario materials, on-site testing, analysis and reporting.

4. Revise scenario and data capture methods

Based upon the outcome of the proof of concept, some revisions may be need to be made to the scenario elements and methods for generating and capturing T&E data.

5. Conduct initial pilot trial with selected MOPs/limited scenario

The goal here is to gather in a cost-effective manner some initial data from segments of the scenario to ensure that everything is working correctly before deploying the more extensive resources required for the main trial. All major forms of probes and data capture tools/methods will be represented. It is assumed that the pilot trial will comprise two days of data collection with morning and afternoon sessions. The personnel requirements for conducting the pilot trial are shown in the following table. As can be seen, three Navy SMEs will be required, of these the RT1 and SWC should be the same individuals for each of the four test trial runs. These role players will require to be additionally trained in the conduct of the trial, especially to act as T&E confederates under some circumstances (e.g. RT1 fails to amplify appropriately, SWC has wrong focus of attention). Further, training will ensure that having been through the same scenario trial on previous occasions that role players not give off inadvertent cues and maintain the same approach and procedures on repeated runs. A different ORO, who is the focus of the trial, will be required for each separate run.

The specific requirements for the pilot trial are outlined in the following table. Where a source is specified as SIM, this means that the information normally provided by that source will be simulated by making it available through T&E staff who follow a script, or in some cases by the GOD.

Information Sources	How Provided	Ops Room Team Players	Personnel	Workstations Real/Sim
Text messages	Pre-scripted/T&E team	SWC	Navy SME	Real
Link 11 (CCS symbology)	Pre-scripted/T&E team	ASWC	Not required	
External net TG	Pre-scripted/T&E team/GOD	ORS	T&E role player	SIM
Internal net Ops Room. General (C&C)	Pre-scripted/T&E team/GOD	EWS	T&E role player	SIM
Internal net team (AAW)	Pre-scripted/ Navy SME role players	CO	GOD	SIM
GCCS	Pre-scripted/T&E team	ORO	Navy SME	Real
Sensor data-air	Pre-scripted/software	RT1	Navy SME	Real
Tracks-surface/sub-surface	Pre-scripted/T&E team	SCS	T&E role player	Real
Stateboards	Pre-scripted/live player	TS	T&E role player	Real
	Pre-scripted	RT2	Not required	
CANEWS	Pre-scripted/T&E team		T&E role player	SIM

Table 5: Summary of logistical requirements for pilot trial

Analyse/report pilot trial

This will involve a full analysis of the data to determine the following:

- the scenario runs according to plan
- scenario events elicit the appropriate responses
- the responses are logged and recorded appropriately
- game players can fulfil the task roles
- the captured data can be analysed and provide the right kind of information for T&E purposes.
- Preliminary estimates of MOP data ranges and variability.

The trial will be reported to the Scientific Authority as a written technical report and an oral briefing.

6. Refine MOPs and scenario

On the basis of the lessons taught from the Pilot trial, modifications will then be made to the scenario and/or MOPs and possibly the T&E software requirements. Since it is unlikely that all potential MOPs can be assessed, given logistical constrains on the experimental design and availability of resources (see below), a selection of the most salient and meaningful MOPs to include in the main study will be made.

7. **Conduct second pilot trial with more complex scenario, if required.**

Depending on the degree of success with the first pilot trial, this step may or may not be necessary before committing the full resources to the main trial. For now it has been included in the schedule as a safeguard.

8. **Analyse second pilot/Refinement of MOPs and scenario**

Same as for items 5 and 6 above.

Note: steps 7 and 8 may not be required if the outcome of the first pilot trial is successful, or if any required modifications are minor in nature and will not require to be formally tested again in NCOT or the ORTT. These steps are seen as providing a conservative estimate of a worst case situation in order to plan for resource allocation and project timeline.

9. **Build and proof scenarios for main trial**

The complete scenario for the main trial is completed based upon the information learned to date. The full scenario is tested and rehearsed using T&E personnel to simulate Navy roles and to drive data as required.

10. **Conduct main trial/ collect baseline data**

This is the formal data collection trial. The scope of the trial is determined by a number of factors relating to data reliability as outlined in a subsequent section. It represents the appropriate level of effort to establish reliable baseline performance data for the major ORO functions.

The availability of the OROs is a major concern for being able to conduct this trial over the consecutive sequence of days proposed. We believe that a minimum of eight data sets be captured, each using a different ORO. Plus we should add an additional ORO for contingencies. Given the limited pool of OROs, obtaining such a large sample at any one time may prove to be difficult An alternate approach that may be considered, if the availability issue cannot be solved, is to consider running four ORO's with each performing two separate sessions. Clearly, in this case different scenarios will need to be prepared for the two different runs, and this will require more preparation and development resources by the T&E team than has been determined in this initial estimate.

11. **Analyse and report main trial**

A significant level of effort will be required to thoroughly review all of the data captured. This will include not only events and responses captured by the software, but also any video or audio records and paper message traffic.

12. **Refine measures and scenarios.**

Prior to the conduct of Trial #2, further refinement will need to be made to the MOPs and scenario to accommodate emphases on different warfare areas. The overhead for preparing, proofing and piloting subsequent trials should be somewhat less than the first trial because of lessons learned and experience gained.

The next table provides an initial approximation of the human resource requirements to accomplish the above. Not included in this table is the proposal to visit the ORTT to review existing scenario records for potential MOP usage (see section 3.2) and to observe an exercise in progress (see 5.7.4). It is estimated that these two activities would require 10 HSI® person days.

The estimates provided below include time for travel to Halifax to conduct the relevant activities and assume that a workstation is not available locally for scenario development. The SWC and

RT1 Navy role players should ideally remain the same throughout the pilot trial and also the main trial, although not necessarily the same individuals on the two occasions. For the main trial we have built in an extra half day cushion to allow an additional session to be added in case of problems arising that result in the loss of a test session. A more comprehensive and detailed estimate of the hours, task allocations and costs to conduct all activities up to the pilot trial has been provided separately to the Scientific Authority.

Task	HSI® staff	Navy SME (normally from HSI® staff)	Simulation facility software developers	Simulation facility support staff	Navy SME role players	Navy SME test participants
1 Generate T&E software requirements Code software	3	1	Unknown level of effort			
2 Generate and encode scenario	5 5	3				
3 Proof of concept						
3a Preparation	2	2		Unknown		
3b Conduct	4	1	2			
3c Analysis/reporting	3 5	5				
4 Revise scenario/methods	3	2				
5a Conduct Pilot (2 days of testing)	9	2 5		Unknown	SWC-2 RT1-2	4 (ORO)
5b Analyse pilot	7	2				
5c Report pilot	8 5	1				
6 Refine MOPs/scenario	5	1				
7* Pilot 2 - 2 days of testing	5	3	1	Unknown	SWC-2 RT1-2	4 (ORO)
8* Pilot 2 Analysis	2	1				
9 Build and proof scenarios for main trial	2	2				
10 Conduct Main trial (assume four days with am/pm sessions plus half day set-up, half spare	20	5		Unknown	SWC-5 RT1-5	8 (ORO)
11 Analyse and report main trial	20	5				
12 Refine measures & scenarios	8	4				

Table 6: Approximate personnel resource requirements for each trial phase (numbers are estimates of person days.) *Note: these steps may not be required.

5.1 Issues Concerning Data Reliability

Given the level of effort required and significant resources deployed to collect baseline data, the T&E team to need to make every effort ensure that reliable data are generated to form a baseline for future comparative purposes. A number of sources can be identified that will affect the variance of collected data and thereby provide constraints on its reliability and generalisability. These constraints are listed below together with an assessment of how they will need to be treated in the T&E trials.

Subject Variability Among ORO's

There are a number of factors that will influence the performance of different OROs. These include range and depth of experience, individual abilities, individual motivation and recent familiarity in doing the ORO functions required in T&E. Variability in each of these domains can seriously widen the confidence limits around mean data. In order to address these issues a combination of selection constraints and choice of appropriate numbers of test subjects is required. Therefore, we suggest that selection be limited to currently active ORO's with a minimum of one year of operational experience.[12] Selection may also include OROs who have been out of operational service for less than one year but have three or more years of prior operational service. A minimum sample size of 8 ORO's will be required for each test trial in order to provide reliable estimates of inter-subject error variance and to obtain sufficient statistical power to detect any performance differences resulting from the TDP (see 5.2 below). Further, it is preferable to use a different sample of subjects for Trial#2, in order to avoid any potential carry-over effects from Trial#1.

A major practical issue that must be considered is that on the East and West Coast combined there is a theoretical maximum of about 45 OROs who are currently in active service, or have had a tour of duty at sea within the last 12 months. However, because of sub-optimum manning levels, this figure is more likely to be closer to 30. Hence, there may be stringent limitations on the availability of appropriate OROs to participate in T&E trials.

Variance Associated with ORO Workload Factors

OROs operate under a variety of levels of workload. Under high levels of load they frequently switch from task to task, spending less time on each than they would if they were underloaded. In order to assess the generality of any improvements in performance that may result from MSDF technology, it will be important to sample from work situations involving different load levels. Further, because OROs may be working at almost optimum performance when underloaded, there may be a ceiling effect on performance that reduces sensitivity in detecting any further improvement in performance due to MSDF. This suggests that devising a test scenario in which ORO's focus only on a subset of critical tasks without any significant workload loading will yield data of potentially minimal value. Thus, it will be necessary to ensure that tasks result in a sufficient level of workload that ORO's are not operating at their maximum capability.

Workload loading for the ORO is not necessarily a homogenous variable that can be characterised by a simple quantitative measure. Workload may vary in terms of the volume of data within a task domain (i.e. ranging from a small number to a large number of air contacts). It may also vary in terms of the complexity of the data within a domain, for example many easy to identify air contacts

[12] Of course, if the Navy is interested in collecting data on how experience affects performance on these tasks, it will be necessary to select two groups of participants who differ in mean years of experience

that are widely dispersed represents a much lighter load than that same number of contacts who are unknown, or enemy and clustered along vectors and altitudes. Further, the number of ongoing tasks across domains will also influence workload, for example when the ORO must co-ordinate responses to potential air, surface and sub-surface threats. The sources of workload variability that may influence ORO performance are outlined in the following table. For within-domain sources, only information relating to the RAP is elaborated, since this is the area of initial priority, and the other domains will follow a similar pattern.

SOURCE OF WORKLOAD	WORKLOAD FACTORS
Within domain	Number of simultaneous contacts Rate of contact Number of unknowns ID ambiguity Track ambiguity Path predictability Number of lost tracks
Between domains	Number of domains (air, surface, subsurface) Differences in volume of data for different domains Number of threats across domains Priorities of threats across domains
Process monitoring	Number of ongoing processes Personnel experience and resource levels Equipment problems Demands from TG Demands from CO

Table 7: Sources of ORO workload

Given that MSDF technology to support process monitoring does not fall within the scope of the COMDAT1 TDP, the focus on workload manipulation for present purposes should be on the other two sources. In order to provide a representative range of workload, it is proposed to have three levels: two levels of workload would be achieved by varying within domain factors and the third by introducing workload from another domain. While it might seem initially appropriate to concentrate on manipulating load factors solely within the air domain (since the TDP will largely centre on air), we believe that this would be an error. Under conditions of high workload in the air domain and in the absence of any other task demands, the ORO would focus exclusively on air warfare. However, this would not be representative of actual operations, where the ORO has always other tasks to timeshare, and would therefore result in some overestimate of ORO performance. However, if we supplement a high within-domain load by a moderate load in another domain, or from other concurrent tasks, the resulting loading will have greater external validity to the operational situation.

We propose that within the overall trial design that **three levels of workload** be sampled as follows:

Moderate workload: The ORO and team are working at a steady but comfortable pace that can be easily sustained, and can cope with the rate of information. The majority of contacts come from the domain of interest, however there is a constant but low level of contact information from other domains.

High workload: The ORO and team are reaching the point of overload; they are working at a high pace that causes some stress if sustained; they can barely cope with the rate of information and some tasks are truncated or dropped; errors may be made. This high information rate is confined to the domain of interest, the contact information from other domains would be sustained at the same level as in the moderate workload condition.

High workload+ extra-domain loading: As above, plus the type and volume of information in the other domains reduces the ORO's capacity to focus largely the air domain. The additional information may involve dealing with potential surface or sub-surface threats.

5.2 Research Design Trade-offs and Sample Size Considerations

One of the major concerns in conducting this form of T&E activity is the constraint imposed by the availability of the facility, support personnel and participants. Unlike an environment specifically designed for research, with a dedicated complement of support personnel and readily available subjects, the primary purpose of NCOT and the ORTT is for Navy training. It seems likely that T&E opportunities will be limited in frequency and duration. As a consequence, the research design must be tightly focussed and sampling procedures must be highly efficient. Trade-offs will have to be made about the extent to which a comprehensive data set can be gathered for establishing a database of baseline ORO performance and to evaluating the effects of the TDP. One way to approach this problem is to consider the number of individual data points that will need to be captured and then work backwards up through the design, to see what is feasible in the likely time to be allotted and personnel (test subjects) to be made available.

The best way to approach this problem is to address issues of the expected magnitude of effects of interest, the anticipated error variance, the acceptable probability of making a Type II error, and the statistical power required. By providing *a priori* ranges of values for these parameters we can readily determine the number of data trials that will be required.

Another factor to be considered in estimating the number of data points required is whether any given MOP will be time based or accuracy based. In practice, accuracy based MOPs require far more trials to achieve the same size of equivalent error variance than response time MOPs. Ten measures of response time per individual will give a reasonable estimate of mean and variance. Whereas ten measures using proportion of items correct can easily be influenced by outlier performance on one or two trials. Further, one would expect that for most ORO tasks relating to situation awareness and communication, training ensures that performance operates at a high level of accuracy. If this is the case, detecting any differences in performance attributable to the MSDF TDP will be difficult to impossible, because of the existing performance ceiling. Therefore, the general focus on evaluation will be to use time-based measures, supplemented by accuracy measures, whenever there are clear instances of tasks that produce consistent errors in performance.[13]

5.2.1 Magnitude of Effects

We have no advance indication of what expectations the Navy may have concerning the effectiveness of MSDF in improving performance. Is a 5% gain of operational significant?- possibly not, unless the task is repeated with high frequency. Performance gains of 10% or more are likely to result in increased efficiency that translates into increasing the ORO's spare capacity. Since we cannot anticipate Navy expectations in this regard, *we recommend that the Scientific*

[13] This general argument does not apply to those MOPs that will require subjective analysis of ORO actions and responses by SMEs, typically for ORO functions involving situation analysis and decision making

Authority in discussions with the Navy should address this issue. Additionally, we could have SMEs review the display concept during prototype development and estimate the magnitude of the changes in performance that might be expected for different tasks. The data generated by this process could then be used to focus the MOP effort. For example, in areas where large changes are expected, or where the changes are too small to be of operational interest, then there would be no need for a large concentrated T&E investment, instead the T&E effort could be better directed to areas where the impact was less certain.

In the interim, we will proceed with determining estimates of required sample size by exploring the implications of effect sizes of 10% and 20%.

5.2.2 Expected Error Variance

Normally estimates of error variance (that are used to guide sampling decisions) are obtained from the relevant literature or pilot studies. In the present circumstances neither of these sources appear to provide any useful guidance. We do not know how long some of the ORO tasks might take, or the variability underlying them. Therefore it seems appropriate to generate a possibility matrix that covers the range of circumstances that may be anticipated to guide the decisions on required sample size.

Based upon our knowledge of the ORO functions, observations of exercises and familiarity with C2 operations in other domains, we propose that the lower range of task completion times of interest is probably 5 seconds. This may seem too short a time to consider from the perspective of achieving meaningful operational increments through MSDF. However, if the tasks that produce this kind of response latency are highly frequent occurrences, small saving in efficiency will accumulate to the point of being operationally significant. At the other end of the range, we can only make a guess as to how long some tasks may take. Arbitrarily we must make some cut off, otherwise if an event produces a typically response that may take 20 minutes to evolve, there will be insufficient time in any one test trial to have adequate repetitions of such events. For working purposes, we have chosen an upper limit of 5 minutes as being the longest response latency that we can effectively deal with. This then gives us a range of potential mean response times between 5 seconds and 5 minutes. We can then interpolate some values between these limits and look at the impact of the range of expected means on sampling requirement by taking into account anticipated variance around these means.

For the purposes of generating some idea of sampling needs, we recommend as a starting point considering standard deviation values that represents 10, 20 or 30% of the mean. These numbers are based upon what might be typically found in complex reaction time or search experiments, although sometimes standard deviations that are 50% of the mean are found. However, such large values impact severely on the ability of the study to be sensitive to differences of interest.

The magnitude of the effect of interest and the error variance can be used to calculate a standardised effects size (**d**) which is defined as (Mean 1 - Mean 2)/ SD. Using this definition, in the standard psychology literature, **d** values of 0.2, 0.5 and 0.8 are regarded as small, medium and large effects, respectively.

Taking our working example of a 5 second RT and differences of interest of 10 or 20%, with estimates of the SD as being 10, 20 or 30% of the mean value, we arrive at the following table of **d** values. As can be seen, these estimates represent extremely large values for d, well beyond what is regarded in the literature as being large. However, we can use the desired magnitude of the difference and the estimates of the variance to estimate the sample size requirements as outlined below, once we have discussed issues of statistical significance and power.

Possible mean value (seconds)	5	
Magnitude of effects	10%	20%
SD as a proportion of the mean		
10%	1	2
20%	0.5	1
30%	0 33	0.67

Table 8: Values of d computed for various assumptions concerning effects sizes of influence for test case RT = 5 sec

5.2.3 Desired Level of Statistical Significance

Typically, research in the social sciences sets the upper limit for accepting that the observed results are due to chance (%)= .05%. This means that five times in a hundred we will incorrectly conclude that when there is no effect, we will say there is one. This level may be too stringent for the kind of exploratory research and investigation that will form the TDP evaluation trial, and to establish a set of baseline performance measures. Here the goals will be to overcome the many sources of error variance that could work to make the study less sensitive to differences of interest, and maximise the chances of detecting any potential change in performance that could be of operational importance. Hence, we recommend a slightly less stringent level for making Type 1 errors by setting %=.1. In practice, this will mean fewer trials will be required to achieve a statistically significant outcome.

5.2.4 Desired Level of Power

Statistical power refers to the ability of the design to establish performance differences of interest, or the odds of confirming a theory correctly. Obviously, one would want this to be as a high as possible within the realms of what is achievable and practical within the constraints of time and effort. Too little power will result in a situation where the study has little chance of detecting significant effects. Too much power will mean that too much data are generated to the point that trivially small effect sizes are detected. In recent years a common, yet arbitrary, choice for a power level is .8

5.2.5 Implications of the Above for Sample Size

Having selected approximate values of the magnitude of the effects we are interested in, the range of anticipated possible mean values (for response time MOPs), the range of error variance and established significance levels and the desired power, we are now in a position to determine the impact of these variables on the required sample size. The following table shows the required sample sizes for the range of values outlined above. The top row provides three different possible values for means, the second row shows the size of difference between means that would be of interest and the body of the table shows the sample sizes that would be required under three different assumptions about the size of the standard deviation. A further assumption is that differences will be assessed using a two-group F-test (analysis of variance).

Possible mean value (seconds)	5		20		40	
Magnitude of effects	**10%**	**20%**	**10%**	**20%**	**10%**	**20%**
SD as a proportion of the mean						
10%	8	6	8	6	8	6
20%	28	8	28	8	28	8
30%	58	16	58	16	58	16

Table 9: Estimates of required sample size to reach required conclusions

Note that whatever we select for the estimate of how long the RT may be, the estimates of sample size all come out to be the same.

To interpret this table let us take a mean baseline response time of 20 seconds, and assume that a difference due to the TDP of 10% would be of interest, this would require a sample size as low as 28 if the SD were about 4 (i.e. 20% of the mean), but a sample size of 58 if the SD were about 6.7 seconds (30% of the mean). The detection of a 20% difference between baseline and TDP would require samples of 8 and 16, respectively for the same variance assumptions. Two, somewhat obvious, but important general principles follow from this, and should be always remembered in considering the design of the study. First, smaller effects are harder to detect and require larger sample sizes. Second, higher error variance reduces the ability to detect effects of interest and also increases sample size needs.

Based upon the above, a reasonable goal for the study would be to try to detect differences of 20% and to keep standard deviations around 20% of the mean value. If this can be achieved then we will need a sample size of 8 participants for each condition (i.e. baseline and baseline plus TDP).

For the present, we should bear the above number in mind in considering the overall demands on the time of ORO participants. Clearly we are faced with a number of unknowns. First, we do not know the extent of the ranges of the independent variables or how many levels of each will be required. Second, we have no estimates of the kinds of performance levels we can expect of the ORO participants. Third, the extent of inter-ORO variability on task performance is unknown. Fourth, we do not know the capability of the system to collect reliable MOPs data. *Consequently, it would seem prudent to not go into detailed design issues beyond Trial 1 at present.* Once this trial, or even the pre-cursor pilot trials to these have been completed, we will be in a safer position to outline the specific requirements for the subsequent trials.

5.3 Requirements for Scenario

The following table outlines the major elements that will comprise the scenario.

Geographical context

- Littoral environment between two large land masses that comprise Nations A and B, approximately within 30 nm each side of ship
- Land masses have mountainous areas coming close to coast that create hills and valleys[14]

Political context

- Nation on left and right land masses are potentially hostile to Canada/Allies and hostile to each other
- Tensions are high between both countries
- There are threats by both sides to embargo international waters to oil exports

Military context

- Halifax class ship operates in a Navy TG comprising, a high value unit, one Iroqois class and two other Halifax class frigates
- A US battlegroup is within 150 nm of the Canadian TG. Intensive carrier based flight schedules are ongoing. These may generate between 10-20 friendlies in the sky at any time (10 for baseline workload, 20 for high workload)
- Nations on left and right each have airfields within 10 nm of coastline
- Nations on left and right conduct regular air training ops that include simulated air combat and bombing runs

Background theatre

- Area is in vicinity of two major commercial air lanes. One is medium-high-level with a/c largely in transit, the other is for a/c that are landing and taking off from an airport located close to one of the military bases
- Local helicopter traffic to and from oil rigs
- Local surface traffic comprising a mix of small vessels (e g media, recreational), commercial traffic (e g fishing vessels, tankers, and bulk cargo carriers), fast patrol vessels belonging to adjacent nations

Op Orders

- Range of tactical area of interest (AOI) is a radius of 125 nm
- No pursuit or engagement of threats

General level of commercial air traffic

- For baseline workload- 30 commercial a/c in AOI - appearing and dropping off at a rate of about 1 every 60 sec
- For high workload - 45 commercial a/c in AOI – appearing and dropping off at a rate of about 1 every 45 sec

General levels of threat/unknown

- For baseline workload- 2-5 fighter bombers at any one time
- For high workload - 7-10 fighter bombers at any one time

Table 10: Scenario Requirements

5.3.1 Approximation of Number of Events per Test Session

In order to make the scenario and associated tasks in building the RAP realistic, we should conform to the kinds of rates of information that might be expected in actual operations as well as an appropriate proportional mix between non-threat and threat events. Thus, we would not expect that all contacts presented during the trial to be contacts of interest, nor will every contact be enhanced by MSDF. It would be unrealistic to have all contacts as unknown/possibly hostile, since this would bias the performance of the ORO and air team in a way that does not normally occur in either real operations, or in training simulations. Instead, the targets of interest (from the point of view of the evaluation of the TDP) have to be embedded within the normal stream of contacts that would be encountered. The question then is what would be an appropriate rate. The practicalities of the design requires a high number of contacts (since we do not care much about the contacts of non-interest), yet realism and the avoidance of bias demands a more moderate rate. As indicated

[14] This particular environment is known to produce track identity problems for radar systems that are primarily designed for open water The map database for NCOT is not able to generate this degree of land mass complexity, hence the behaviour of radar under these circumstances will be simulated using real-time control by the T&E team of the game entities underlying the radar tracks

above, we suggest that non-threat contacts occur at a rate of one per 60 seconds up to a maximum of 30 on the screen at any one time for the moderate workload condition. They also drop off the screen at the same rate. For the high workload conditions contacts would occur every 45 seconds up to a maximum of 45 on the screen. For a two-hour trial, there would be a total of 120 and 160 contacts for the moderate and high workload levels respectively. We suggest that we superimpose on this a rate of between 25-35% for the contacts of interest as a percentage of the overall number of contacts. This will mean about 30-40 (moderate workload) or about 50-60 (high workload) trials that will be available for MOP data collection. Given that a sample size of about 10 repeated trials for each measure will be required, this suggests that the design will handle between about 4-6 different types of events involving different behaviours by contacts of interest.

In addition to accommodating the main session in which data will be collected, the design will also need to provide an initial period of training for the participants. It is suggested that 30-45 minutes at the start of the test session be allocated to this.

5.3.2 General Structure of Test Trial and Course of Events

1. **Pre-watch scenario build-up**. Prior to the start of data collection, the background picture is built up by the T&E team. This will probably take 10-15 minutes.
2. **Watch handover:** ORO is briefed - about 5 minutes
3. **Phase in**: Routine watch activities to allow a period of relative peace and quiet for ORO to become aware, settle, and for T&E team to collect routine new contact cycle data. Air situation =25 commercial, 4 friendly in AOI (Note in high workload/across domain condition will also need starting scenario for MSP, MSubP)- about 10 mins
4. **Main scenario events:** this will take about 90 minutes and have the following features.

- Commercial traffic added and dropped continuously
- Adjacent nations A and B are conducting ongoing air ops with take-offs every 5-10 mins, circuits and occasional flights towards each other.
- Unknown/possible threat air ops involve a/c doing some (not all) of the following event types[15]
  - simulated bombing runs on range
  - flying through hills and valleys
  - flying in circuits that progressively get closer to the TG
  - close formation and changing formation
  - attempting radar stealth
  - two a/c on same course/speed in close formation- one at constant altitude the other descending/ascending
  - two a/c that converge/run together for a time, then diverge
  - two a/c in close formation on a general heading to overfly the ship; a/c criss-cross with increasing frequency as they get closer
  - three a/c in close formation with one breaking off and criss-crossing the path of the remaining two

[15] These are based in part upon the a/c flying patterns used for the ASCACT trials and are believed to create problems for current sensor systems

- missile separation from one a/c (i.e. when a new air target appears suddenly and close, a pop-up target)

Given limitations outlined in the previous paragraph in terms of number of trials/events of interest, some prioritisation of these event types will need to be made.

5.4 Methods for Data Capture

5.4.1 Software Capture

The software will be required to capture and record the time (preferably to nearest .5 sec) of specific trigger events during the course of the scenario[16], the full details of these will be worked out during the detailed scenario planning and initial proof of concept trials. The events to be captured will include: specific key presses by the ORO and other Ops Room team members (where applicable) and selected key presses from the T&E team console(s).

5.4.2 Video and Audio Capture/Analysis

The system will be required to capture and record the time (to nearest 1/10 sec) all audio and text based comms including their content and the recipient/sender. A complete audio record will be maintained on tape (or equivalent) of all comms, including those involving the Ops Room team and those between the Ops Room team and the T&E team. ORO comms that are direct and do not go through a network will also need to be captured. This may require a separate microphone on the ORO attached to an audio tape recorder.

The system will be required to capture screen contents from the ORO CCS on a time base accurate to 1/10 sec. The specific requirements will be worked out during the detailed scenario planning and initial proof of concept trials. The resolution of the screen capture shall be sufficient to be able to discriminate track details, symbology and all text.

5.4.3 T&E Probes

The system will allow the T&E team a capability to send from a remote terminal a message to the ORO in one of three formats: audio, screen based text, or paper-based text (hand delivered). A time stamp at the moment of issuance of the audio and screen based messages will need to be recorded. Any response to the message required by the ORO that will use the CCS or audio system will also be time stamped.

The system will allow the scenario to be stopped (and the tactical data on the ORO screen may be required to temporarily eliminated) to allow the T&E team to conduct probes of the OROs knowledge of screen content. These probes will either be by audio communication or in the form of a message sent to the ORO console from the T&E console. The system will allow the scenario to restart with full data intact with no scenario elapsed time at the conclusion of the T&E probe (likely within 3-4 minutes maximum).

---

[16] If the existing software cannot be modified to perform such timing, then the intervals of interest will have to be determined after the event by having the T&E team scrutinise and analyse the playback record This will create a considerable additional burden of analysis that has not been factored into the interim estimates of personnel resource requirements

5.4.4 SITREPS

To meet some T&E purposes, the ORO may be required to provide a brief SITREP. Options for the SITREP include either a verbal report (recorded on audio tape or by the software) or a written "fill in the blanks" report. We favour the former as it is more natural, less time consuming and does not provide the kind of contextual prompts available in the written format. The software will need to record the time of the request and time of completion of a SITREP and the screen contents of the ORO workstation at the time of the request (or any other screens that the ORO uses in order to comply with the request).

5.4.5 SME Real Time Observation

In order to evaluate some of the more complex ORO behaviours, we recommend the use of an experienced ORO stationed in close proximity to allow observation of the ORO test participant, much in the same way as training is now conducted. The role of the SME will be to evaluate ORO performance along prescribed criteria to be developed during the scenario construction phase. The observer will be provided with a workstation that can display either the same content as that of the ORO or any other information currently capable of being generated by the scenario. The workstation will also have the capability to capture selected key presses of the SME and allow messages to be created and notes to be recorded.

5.5 Data collection and Management Tools

It follows from the above that the following types of data are to be collected and that each will have its own requirements for management. In general, it would be desirable if all forms of critical data for analysis are provided in a medium or format that allows them to be taken off-site for analysis on a standard Microsoft Windows based platform (assuming that there are no security issues).

5.5.1 Response Time Data

These data are collected by the simulation software using prescribed trigger events or flags initiated by the T&E team. An underlying timebase with a resolution of 0.1 sec from the start of the trial must be maintained in order to mark events. Any individual response time is terminated by an event determined during scenario construction.

Response times (RTs) may also be required for events that will require a verbal response by the ORO, in these cases, the logging of such responses will also require a running timebase. (See also audio communication below).

The software will allow RTs associated with events to be coded in such a manner that they can be quickly retrieved and organised after a test session is complete. It would be preferable if this retrieval process could be expedited on the same day as the trial and be handled by the T&E team, rather than requiring the data to be sent to the software contractor for extraction. The data extracted by this process will be imported into an Excel spreadsheet where it may be organised appropriately for analysis.

5.5.2 Accuracy Data

Accuracy data may relate to several different kinds of information. These include the accuracy in knowing and responding to screen content (e.g. location and meaning of symbology, tactical data of importance), accuracy in making tactical estimates (e.g. closest point of approach), accuracy in communicating, accuracy in situation assessment and decision making. Given the variety of the information content in all of these possibilities and the different modes of response, a single prescription for the collection and management of accuracy data cannot be provided. In some cases the information will be retrievable from ORO key presses or SSD screen-dumps, in other cases from text or oral communications, in other cases from responses to T&E probes (either in real time or if the scenario is temporarily frozen), in other cases it may have to be provided by the observer SME in real time or post event scenario playback.

The process of determining the accuracy of responses will be largely a manual, post-event analysis conducted by the T&E team on the basis of replaying the appropriate segments of the scenario, examining the events of interest and manually recording the responses that were made. In other cases, accuracy data will be provided from the SME observer in either interval data form or as a behavioural rating on a pre-determined scale. Whatever method is adopted, the resulting data will also be entered into an Excel spreadsheet.

It follows from the above, that the ability to replay, fast forward, slow down and halt a scenario record is a pre-requisite for this type of analysis. This requirement applies not only to data captured by the software but also to audio and video data.

5.5.3 Audio and Video Data

Two forms of video data will be captured - screen contents on workstations of interest and a video record of selected players. The screen content data will be captured by the system software and should be amenable to post-event analysis by replay on a workstation. The replay should allow a running timebase to be displayed at all times. An ability to extract single images of screen dumps during replay would be desirable. These should be in a graphics file format that allows them to be viewed with standard commercial graphics software. A JPEG format would be preferred to maintain file sizes at a manageable level. These data should be capable of being exported to other systems outside the T&E environment.

The video record of selected team players will be capable of being played back on standard commercially available equipment and will provide an on-screen running timebase.

The audio record will comprise all net-based communications and those captured directly from the ORO's microphone, that do not go through the standard comm channels. The audio record should be capable of being married to the video record where appropriate.

Audio and video records will be managed by maintaining a log and database. Copies of all critical records will be made for back-up purposes.

5.5.4 Observer SME Ratings

To review, SME ratings may be recorded in real time during the course of a trial or may be generated during post-trial scenario replay. In both cases, the data will be entered into an Excel spreadsheet in a format that allows events and their associated evaluations to be correlated. All subsequent analysis of the ratings will be maintained within the spreadsheet and associated with other MOP data where appropriate.

5.6 Data Analysis

The tools for data analysis will either be the resident statistical analysis functions in Excel, or where these are insufficient, the data will be exported to SPSS for analysis.

Data for each measure will be collapsed over samples and a representative statistic applied for central tendency. This will normally be the mean, but could be the median for data samples that are highly skewed (e.g. response times). Confidence limits around each mean will be calculated.

To compare baseline with baseline+TDP performance, an overall multivariate analysis of variance (MANOVA) will be used that includes all relevant MOPs. Subsequently, univariate ANOVA, t-tests and planned comparisons will be used to assess any differences in performance on individual MOPs. Magnitude of significant effects will be expressed as a percentage increase or decrease in performance. Should high data variance occur across subjects that masks the identification of significant effects, statistical tests might be performed either on difference scores (TDP-baseline), or by using a non-parametric sign test (e.g. TDP-baseline is expressed simply as a plus or minus depending upon the direction of the performance difference).

5.7 Constraints/Limitations/Risks

5.7.1 Operational Realism

The emphasis on designing the T&E trial has been to maximise the opportunities to collect robust and reliable data. As a result, some operational realism may be lost, for a number of reasons.

- The workload associated with the processing of unknown tracks and contacts of interest may be higher than that experienced operationally or in training. As such, when conducting the first trial (i.e. centred on the RAP) the ORO may focus more on air operations than would be normally the case.
- Many of the other factors that contribute to workload and the distractions that occur under normal operational circumstances will be absent.
- In real operations there may be considerable lulls in the level of background air traffic and in the rate of appearance of contacts of interest. Thus, workload may at times be somewhat low and then followed by a period of more intense activity. The constraints resulting from maximising the opportunities to collect data of primary interest means that such lulls cannot be afforded in the scenario event sequence.
- Although T&E participants will be thoroughly briefed and have a warm up period, by comparison with operational reality they will be dropped "cold" into a busy scenario with which they may have little familiarity. This is unlike the operational context where accumulated experience over previous watches will serve to guide and influence performance on the current watch.
- In reality, performance by the ORO or other relevant members of the Ops Room team is shaped by the cumulative experience of working together as a team. However, in the T&E trials such team cohesion will be absent, as the ORO will largely be working with individuals for the first time.

The first two of these factors suggest that, the T&E trial may overestimate ORO performance levels compared with what might be expected in operational conditions – i e. performance would be worse during operations. The remaining factors may lead to an underestimate of operational performance in the test trial. Nevertheless, relative differences between baseline performance and performance with any COMDAT upgrades should become apparent.

5.7.2 Generalisability

The two major issues of generalisability are: (i) to other OROs beyond the T&E sample and (ii) to operational contexts. With respect to the former, as long as the required number of T&E participants can be obtained and given the small size of the overall ORO population, then adequate generalisability should result. Certainly generalisability will be much higher than in typical research endeavours where the test sample represents a very small proportion of the underlying population.

Generalisability to the operational context is subject to the issues outlined above with respect to operational realism. Increasing such operational generalisability may be addressed by an incremental approach to T&E in which the scenario may be repeated under increasingly higher levels of realism and involving greater contextual complexity. Thus, while initial evaluation may be conducted in NCOT, subsequent trials in the ORTT and then at sea would serve to provide data that confirm or disconfirm the operational generality of any findings. At present the generalisability of the data to other mission types such as assistance of civilian authorities and drug and smuggling interdiction cannot be estimated. By the same token it is not clear that MSDF technology, at least as exemplified by the initial TDP, is designed to assist decision making in these other mission contexts.

5.7.3 Risks: Implementation of MSDF Technology

The greatest risk that can be anticipated at the present time concerns how the MSDF TDP will be integrated into the existing simulation environments. We believe that Lockheed Martin intends initially to integrate the technology into the CSTC environment. As we have indicated previously, there appear to be significant limitations in the CSTC with respect to supporting the additional requirements for T&E, and collecting data with the required reliability and precision. Many of the planned measures outlined above would require a significant increase in capability of the CSTC simulation software to capture the required data. Even if the environment could be adapted to better accommodate the needs of T&E and the data could be captured appropriately, we would be faced with the situation of comparing baseline data collected in NCOT with MSDF trial data collected in the CSTC. Significant differences in a variety of variables across these two environments may either make comparisons impossible to interpret or reduce the sensitivity of the design to capture performance differences of interest.

A pragmatic approach to solving this problem might be to observe how MSDF impacts upon the information processed by the RT1 in terms of radar and track data, and then simulate these effects in NCOT for those T&E trials designed to evaluate the effects of MSDF. The practicality and feasibility of this approach of course remains to be evaluated.

5.7.4 Risks: the Need to Gain Direct Familiarity with Operational Functions

The accumulated knowledge of the T&E team to date has been based upon training exercises in the CSTC, evaluation of the Ops Room deficiencies, several days of scenario based interviews with OROs and other command team members, and information provided by Navy SMEs within the team. The MOPs recommended above are in many cases for behaviours that have not been closely observed but only inferred from verbal descriptions. Clearly what is required to round out the knowledge of the team is an opportunity for direct, sustained observation of actual Ops Room functions in progress by the T&E team. This will mitigate the risk of pursuing inappropriate MOPs and allow the T&E team to make a more informed final decision on which MOPs to include in testing

Now that the ORTT is in place and fully functional, we can believe that this can be readily and easily achieved. Therefore, *we recommend that before conducting the pilot trial, the T&E team either visit the ORTT while a full training exercise is in progress, or review existing records of training using the ORTT playback capability.* The goal of this will be to observe each of the major Ops Room functions in execution, with a view to validating the proposed MOPs and possibly uncovering critical tasks that may have been overlooked, for which MOPs should be developed. Of the two options proposed, the playback of existing scenarios may have several advantages.

First, the T&E team would not "get in the way" of live exercises or simulator training. Second, the team can stop and replay the scenario for better analysis. Third, a Navy SME can be used to assist in the analysis without any of the time pressure that occurs in live scenarios. Fourth, it may be possible to review records of several different teams to determine the degree to which processes are standardised or vary.

The logistics of this would involve three members of the T&E team (including one Navy SME) conducting a three day visit to the ORTT. If the playback review of previous records option is chosen then the analyses would be conducted at times when the main ORTT simulation facility is not being used by the Navy.

5.7.5 Risks: Lack of Data on Using Training Simulators to Measure Operational Functions for T&E Purposes

Worldwide, there is very limited experience in using training simulators for T&E purposes, and human performance data from operations or simulators is limited. The T&E demands for precision, reliability and repeatability may not have been the anticipated in the design of Navy simulation suites whose first priority is training. This is evidenced by the difficulty encountered in trying to retrofit measurement technology in the CSTC and the resulting insufficiency of the data produced to meet the more rigorous needs of T&E. Further, in the case of the NCOT facility, the proposed T&E program may push its envelop of capabilities and will be the first in-depth attempt to run team-interactive scenarios with multiple interacting workstations.

5.7.6 Risks: Availability of OROs as Test Subjects

The test plan is predicated on the assumption that sufficient OROs will be available to allow multiple test sessions in order to gather reliable data and establish confidence in estimating effects of interest. However, the reality that must be faced is that potentially fewer OROs could be made available to meet the needs of the program. If this were the case, then an appropriate approach would be to collect more data (in terms of replication of events) for each ORO. To accomplish this an additional scenario would need to be built with sufficient variation in events and actions to reduce carry over effects and minimise the risk of anticipatory responses. The project plan of resource requirements does not presently include this potential requirement.

6. Summary

The major tasks outlined above have been to identify appropriate MOPs, to consider how they may be implemented in a T&E scenario, to construct an initial overall T&E plan, to consider test design implications and to specify logistical and resource requirements for testing.

In reviewing the MOPs selected, it should be noted that the focus has been largely on those functions that impact upon the ORO's situation awareness process of detection, integration and comprehension of information from various aspects of the tactical pictures to support command decision making. As such, these areas stand to be most impacted by the short to medium term MSDF technology developments of COMDAT1. For many of the other principle functions of the ORO concerning people and process management, the shape and potential viability of the MSDF technology that may enhance such processes is unknown. Therefore, no attempt has been made to consider appropriate MOPs for these functions at this stage.

A large number of MOPs have been identified, probably more than can be practically implemented in the main T&E trials to collect baseline data. Those measures that turn out to be impractical, unreliable or requiring undue overhead for the return, will probably need to be set aside, as lessons are learned in proof of concept and pilot trials. However, a further consideration that must be taken into account when evaluating whether to retain an MOP concerns its diagnosticity or overall effectiveness. The fact that a specific MOP may be accurately and reliably measured does not address the issue of its overall utility. For example, if a particular sub-process in the detect-to-classify sequence can be accurately measured, little useful information will have been gained if it is found to contribute to say less than 5% of the time required for the overall function. A priori, in the absence of detailed information flow process diagrams for these tasks with associated network simulations of process times, or being given access to operational performance data sets such as those collected by the Maritime Warfare Centre, we cannot provide guidance as to which MOPs are likely to account for the majority of the variance associated with effectiveness. Thus, it appears likely that as data are collected during the pilot and early main trials, evidence will accumulate as to which key MOPs will become the focus for optimum human-system performance description and analysis.

Finally, it should be noted that the CTA and subsequent validation provided an initial overview and framework for understanding the work of the ORO based upon a particular scenario structure. As we continue to understand the role of the ORO under a variety of operational contexts in real time, there will probably be a need to refine, augment and update the original analysis. As will also be the case to reflect ongoing changes in Navy command and control concepts, terminology and Ops Room resourcing strategies

THIS PAGE INTENTIONALLY BLANK

Glossary

AAW	Anti-Air Warfare
AWW	Above-Water Warfare
A/C	Aircraft
AOI	Area Of Interest
SWC	Sensor Weapons Controller (also Surface Warfare Commander)
CO	Commanding Officer
COP	Common Operational Picture
CPA	Closest Point of Approach
GCCS	Global Command and Control System
LAP	Local Area Picture
MCOIN	Military Command Operational Information System
MOP(s)	Measure(s) of Performance
MSDF	Multi-Sensor Data Fusion
MSP*	Maritime Surface Picture
MSubP*	Maritime Sub-surface Picture
MTP	Maritime Tactical Picture
NCOT	Naval Combat Operator Trainer
OPGEN	General instructions from the Tactical Commanding Officer to Commanders
OPTASK	Operational instructions from the appropriate commander that detail the conduct of operations (note: there will be a number of different OPTASKs for different warfare areas, such as air, surface and subsurface).
ORO	Operations Room Officer
ORTT	Operations Room Team Trainer
PU	Participating Unit
RAP	Recognised Air Picture
RLP	Recognised Land Picture
RMP	Recognised Maritime Picture
SITREP	Situation Report
SWC	Sensor Weapons Controller (also Surface Warfare Commander)
TDP	Technology Demonstrator Project
T&E	Test and Evaluation
TG	Task Group
WAP	Wide Area Picture
WD	Warfare Director

*Note these are abbreviations coined for present purposes and may not represent current Navy acronyms

THIS PAGE INTENTIONALLY BLANK

References

1. Matthews, M.L., Webb, R.D.G. and Bryant, D. Cognitive Task Analysis of the Halifax Class Operations Room Officer. DCIEM Report # . March 1999.

2. Matthews, M.L. and Webb, R.D G. Validation of the Cognitive Task Analysis of the Halifax Class Operations Room Officer. DCIEM Report # 2000.

3. Matthews, M.L., Webb, R.D.G. and McCann, C. A Framework for of Military Command and Control Systems. DCIEM Report#CR-1997-024.

4. COMDAT I Technology Demonstrator. Statement of Work. DREA 2001.

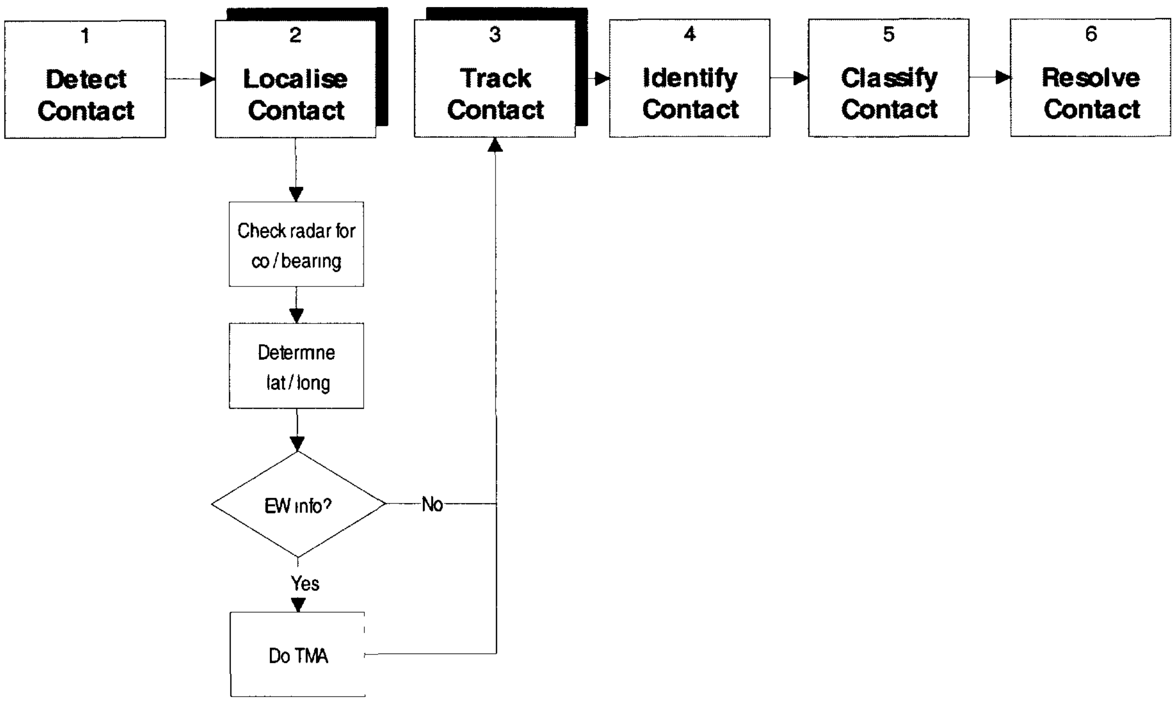
5. COMDAT I Technology Demonstrator. System Concept. DREA 2001

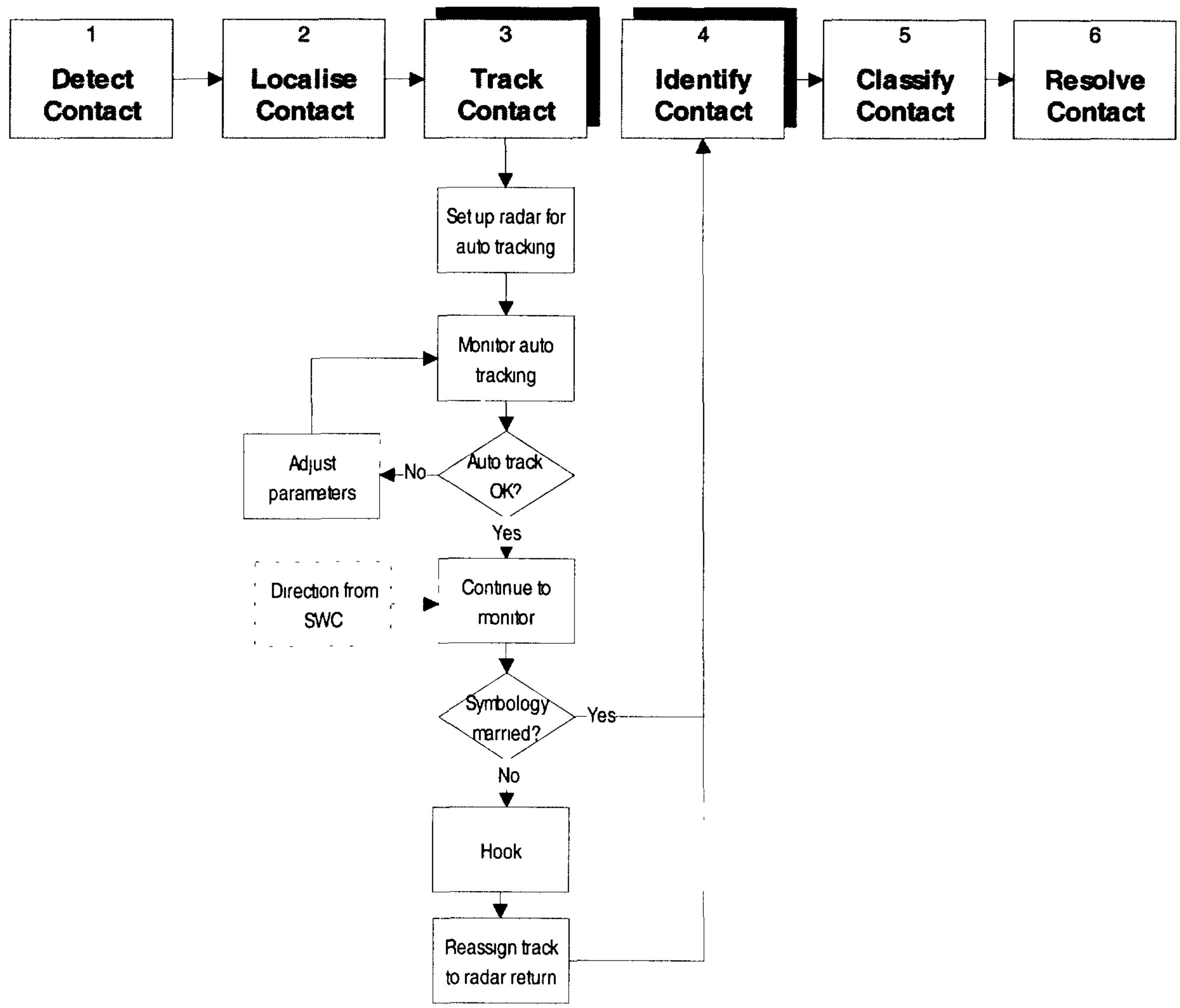
6. The Naval Command and Control Way Ahead: A Blueprint for 2010.

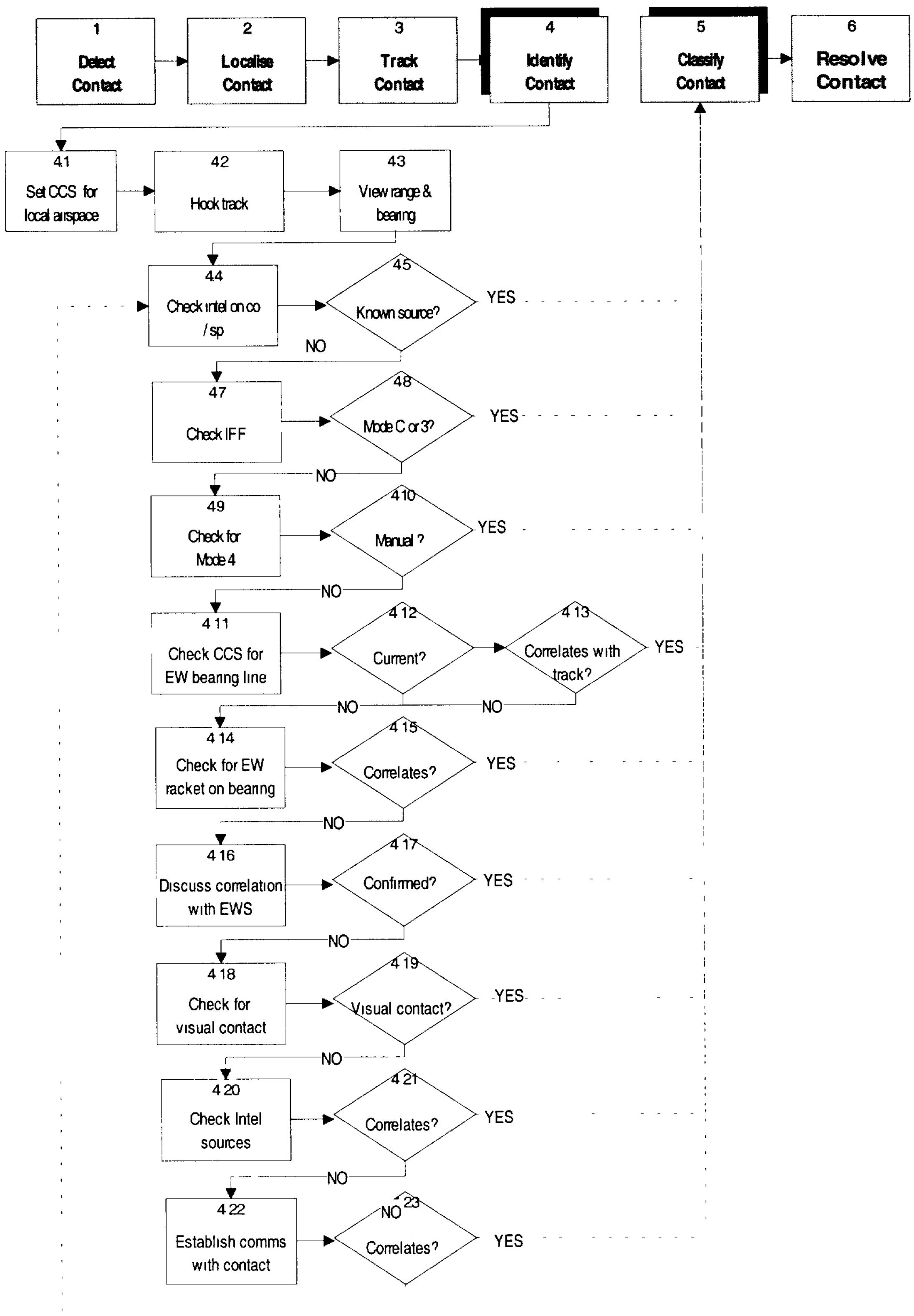
7. AUS-CAN-NZ-UK-US (1997).Handbook 5: Guidelines for Maritime Information Management (version 2.1).Command, Control and Communications Committee.

THIS PAGE INTENTIONALLY BLANK

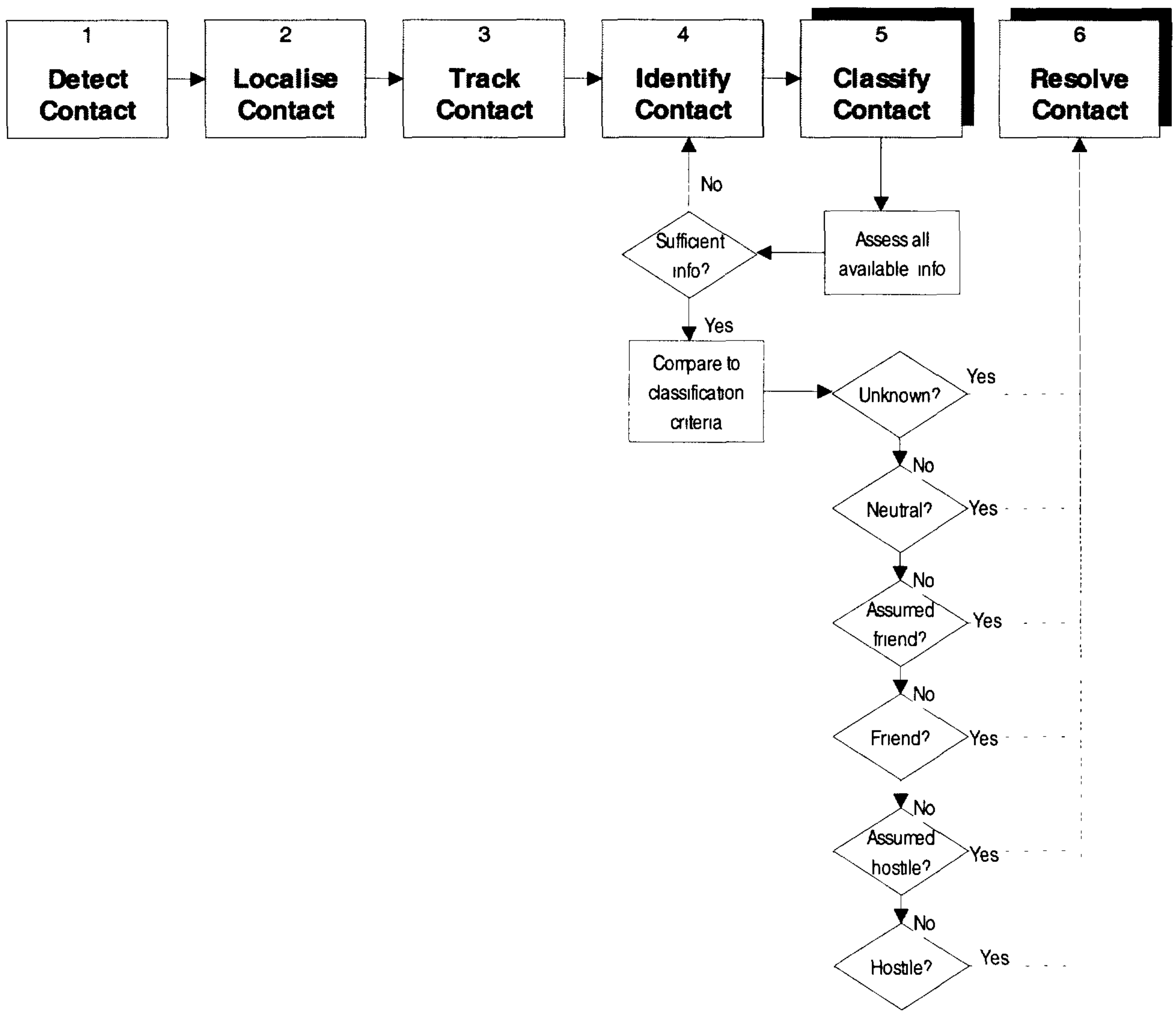
Annex A: Information Flow Diagrams of the Detect-to-Resolve-Cycle


1
Detect Contact
2
Localise Contact
3
Track Contact
4
Identify Contact
5
Classify Contact
6
Resolve Contact
Check radar for co / bearing
Determne lat / long
EW info?
No
Yes
Do TMA


1
Detect Contact
2
Localise Contact
3
Track Contact
4
Identify Contact
5
Classify Contact
6
Resolve Contact
Set up radar for auto tracking
Monitor auto tracking
Adjust parameters
No
Auto track OK?
Yes
Direction from SWC
Continue to monitor
Symbology married?
Yes
No
Hook
Reassign track to radar return



1
Detect Contact
2
Localise Contact
3
Track Contact
4
Identify Contact
5
Classify Contact
6
Resolve Contact
41
Set CCS for local airspace
42
Hook track
43
View range & bearing
44
Check intel on co / sp
45
Known source?
YES
NO
47
Check IFF
48
Mode C or 3?
YES
NO
49
Check for Mode 4
410
Manual ?
YES
NO
4 11
Check CCS for EW bearing line
4 12
Current?
4 13
Correlates with track?
YES
NO
NO
4 14
Check for EW racket on bearing
4 15
Correlates?
YES
NO
4 16
Discuss correlation with EWS
4 17
Confirmed?
YES
NO
4 18
Check for visual contact
4 19
Visual contact?
YES
NO
4 20
Check Intel sources
4 21
Correlates?
YES
NO
4 22
Establish comms with contact
4 23
Correlates?
YES
NO


1
Detect Contact
2
Localise Contact
3
Track Contact
4
Identify Contact
5
Classify Contact
6
Resolve Contact
No
Sufficient info?
Assess all available info
Yes
Compare to classification criteria
Unknown?
Yes
No
Neutral?
Yes
No
Assumed friend?
Yes
No
Friend?
Yes
No
Assumed hostile?
Yes
No
Hostile?
Yes

THIS PAGE INTENTIONALLY BLANK

Annex B:
Follow-up Evaluation of the NCOT Facility

Notes from the HSI® visit to NCOT November 21/22, 2001 concerning MOPS and Test and Evaluation needs for COMDAT1 Evaluation.

1. Sufficient workstations can be made available to meet T&E Needs.

2. Workstations can be configured and networked to simulate a team comprising the RT1, SWC and ORO.

3. Workstations are available and can be configured for the T&E team to monitor the workstations of the Ops Room team.

4. Workstations are available and can be configured for the T&E team to manipulate scenario events.

5. Workstations screens and all communications can be captured and stored on a hard disk for later replay. The capacity limitation has not been tested to date. Audio data are captured to a PC file, video data to the HP workstation and is limited by hard disk size. Hard disks cannot be swapped during a session to enhance capacity. The current video limit is thought to be about one hour with the existing size of disks. This needs to be more rigorously tested with a view to determining the actual limit, what trade-offs in captured functions may have to be made and whether a larger disk will solve the problems for T&E data collection.

6. Flags for T&E purposes are also on DND's wish list but MDA has no current plans for implementing these. In fact, the COTS proprietary supplier is not interested in doing it. As to timings, this might be possible to refine the software because MDA controls some aspects of that. The best approach would be to come up with a short wish list of timing start and stop points for T&E purposes and discuss informally with MDA, who could probably program a menu for each, provided they can be uniquely defined.

7. Audio / video playback: the record system uses proprietary COTS, thus information cannot be played back either on any other machine (e.g. off site at DCIEM or HSI®) unless a licensed system is available with installed software.

8. Need to explore with DCIEM/MDA the cost and logistics of acquiring a system that will run NCOT software and the requirements for installation and maintenance.

9. Radar tracks cannot be simulated on the CCS, hence there is a need for an actor to play the role of the RT1 to perform the task of creating tracks. Similarly, an actor will need to be present to perform the role of the SWC.

10. There are no existing scenarios that can be suitably adapted for T&E purposes. However, the library of already created game entities (e.g. aircraft, ships and their associated attributes and kinematics) can be used for scenario building and hence these will not need to be created from scratch. A brief review of these entities suggests that they are sufficiently diverse in type and number to satisfy the scenario building requirements for T&E.

11. Existing land mass geographical maps are suitable for the purposes of T&E. However, they represent a flat two dimensional surface only, hence desirable elements of the operating environment such as coastal hills and valleys cannot be simulated in terms of their effect on the radar picture. This will mean that the loss of radar signal and signal degradation that would normally occur from air contacts moving in such a region will have to be simulated in real time by a member of the T&E team.

12. In the current software build, the behaviour during the scenario of pre-programmed events is unreliable. The next build scheduled for delivery in January is supposed to rectify this. Currently, there is a lack of trust that the software will indeed handle pre-programmed events in the required manner. Therefore it has been recommended that real time control by the T&E team of some events can be anticipated. This will create an additional administrative overhead in running the scenarios and will require the development of a careful, detailed script that is well rehearsed with highly trained personnel.

13. T&E personnel are able to control the behaviour of entities in real time by making them active or inactive and moving them to new geographical positions while inactive.

14. The facility appears capable of being able to reproduce all of the kinds of events anticipated in the initial scenario plan. Some can be readily implemented, others will require some workarounds.

15. After developing the master events list and scenario schedule, the implementation will require some iteration in testing and refinement. This can only be conducted on an NCOT workstation. The skills to conduct this can be readily acquired by the T&E team.

16. The scenario building tools available are not very user-friendly and must be compensated for by spending more time in paper planning beforehand. The inaccuracy of the system in driving entities can also be compensated for by spending more time during the programming stage - i.e. playing it over and over until the entities do what you want. On the positive side, after developing a good background scenario, the work required to complete a whole scenario by inputting unique foreground events is minimised.

17. The facility is adaptable to adding an ancillary audio and video recording capability for T&E purposes.

DOCUMENT CONTROL DATA SHEET

1a PERFORMING AGENCY Humansystems Inc , 111 Farquhar St, 2nd floor, Guelph ON, N1H 3N4	2 SECURITY CLASSIFICATION UNCLASSIFIED Unlimited distribution -
1b PUBLISHING AGENCY DRDC Toronto	

3 TITLE

(U) Assessing the impact of Multi-Sensor Data fusionon Command and control Operations in the Halifax Class Frigate: Recommendations for measures for performance and detailed test plan

4 AUTHORS

M.L. Matthews; R.D.G. Webb; A.R. Keeble

5 DATE OF PUBLICATION February 26 , 2002	6 NO OF PAGES 69

7 DESCRIPTIVE NOTES

8 SPONSORING/MONITORING/CONTRACTING/TASKING AGENCY
Sponsoring Agency
Monitoring Agency
Contracting Agency
Tasking Agency

9 ORIGINATORS DOCUMENT NO Contract Report CR 2002-018	10 CONTRACT GRANT AND/OR PROJECT NO. W7711-7-7404/01-SV	11 OTHER DOCUMENT NOS

12 DOCUMENT RELEASABILITY

Unlimited distribution

13 DOCUMENT ANNOUNCEMENT

Unlimited announcement

14 ABSTRACT

(U) This report provides a comprehensive Test and Evaluation (T&E) plan for evaluating the impact of future command decision aid technologies (COMDAT) on human and operational performance in the Halifax class operations room. The plan provides details of an incremental approach to T&E that will generate reliable and valid performance data for a range of critical tasks performed by the Ops Room team, with a general focus on the role of ORO. Elements of the plan include: the number and types of test trials, scenario development and implementation, requirements for personnel, logistics and facilities, details of the research design and a comprehensive list of measures of performance and the means by which data will be captured and analysed.

(U) Le présent rapport comprend un plan d'essais et d'évaluation exhaustif pour évaluer l'impact des technologies d'aide aux décisions de commandement (COMDAT) futures sur le rendement humain et opérationnel dans la salle des opérations des bâtiments de la classe Halifax. Le plan présente de façon détaillée une approche progressive aux essais et à l'évaluation qui permettra d'obtenir des données fiables et valides sur le rendement de l'équipe de la salle des opérations dans le cadre d'un éventail de tâches importantes, tout en portant une attention particulière au rôle de l'officier de la salle des opérations (OSO). Le plan comprend : le nombre et les types d'essais, l'élaboration et la mise en oeuvre de scénarios, les besoins en matière de personnel, de logistique et d'installations, des données sur la méthodologie de recherche, une liste exhaustive des mesures du rendement, ainsi que les moyens de collecte et d'analyse des données.

15 KEYWORDS, DESCRIPTORS or IDENTIFIERS

(U) decision aid technologies; COMDAT; Operations Room Officer; data fusion; measures of performance; test and evaluation